ABSTRACT

Title of dissertation: A New Deep-Neural-Network–Based Missing Transverse Momentum Estimator, and its Application to W Recoil

Yongbin Feng
Doctor of Philosophy, 2020

Dissertation directed by: Associate Professor Alberto Belloni
Department of Physics

This dissertation presents the first Deep-Neural-Network–based missing transverse momentum ($p_{\mathrm{T}}^{\mathrm{miss}}$) estimator, called "DeepMET". It utilizes all reconstructed particles in an event as input, and assigns an individual weight to each of them. The DeepMET estimator is the negative of the vector sum of the weighted transverse momenta of all input particles. Compared with the $p_{\mathrm{T}}^{\mathrm{miss}}$ estimators currently utilized by the CMS Collaboration, DeepMET is found to improve the $p_{\mathrm{T}}^{\mathrm{miss}}$ resolution by 10-20%, and is more resilient towards the effect of additional proton-proton interactions accompanying the interaction of interest. DeepMET is demonstrated to improve the resolution on the recoil measurement of the W boson and reduce the systematic uncertainties on the W mass measurement by a large fraction compared with other $p_{\mathrm{T}}^{\mathrm{miss}}$ estimators.

A New Deep-Neural-Network–Based Missing Transverse Momentum Estimator, and its Application to W Recoil

by

Yongbin Feng

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2020

Advisory Committee:
Professor Alberto Belloni, Chair/Advisor
Professor Sarah Eno
Professor Manuel Franco Sevilla
Professor Anson Hook
Professor Timothy Koeth
Professor Massimo Ricotti, Dean's Representative

Dedication

To my parents.

Acknowledgments

I owe my gratitude to all the people who have made this thesis possible and because of whom my life as a physics graduate student has been one that I will cherish forever.

First of all I would like to thank my advisor, Prof. Alberto Belloni, for guiding me through this challenging but wonderful journey. He has always made himself available for help and advice and I can not remember how many times I interrupted his weekends. He is always nice and patient, no matter how silly my question is. He teaches me how to think, speak, read, write, and code as a professional particle physicist. Without his encouragement all the way round, I can not make it to this step.

I would also like to thank Prof. Sarah Eno for teaching me how to work on physics analyses, how to solve hard problems step by step, and how to behave in a large collaboration. The emerging jet search is my first physics analysis and working with Prof. Eno is a wonderful experience I will never forget. I must thank Prof. Nick Hadley as well for his guidance, comments, and help during the meetings and chats. I especially thank Prof. Tom Ferbel for reading and commenting on the thesis draft.

Many thanks to my committee members Prof. Manuel Franco Sevilla, Prof. Anson Hook, Prof. Timothy Koeth, and Prof. Massimo Ricotti, for serving on my dissertation committee and providing valuable comments.

I am grateful to Maryland postdocs I have worked with: Dr. Geng-Yuan Jeng, Dr. Josh Kunkle, and Dr. Markus Seidel. Dr. Jeng taught me how to work in a

hardware lab, to take careful scintillator measurements, and to do physics analyses. Dr. Kunkle provided enormous help when I moved from Maryland to CERN, introducing me to the new environment. He taught me how to build a beautiful analysis framework and how to write elegant code. Having the chance to work with Dr. Seidel is really great and fruitful. He is knowledgeable and experienced, teaches me how to do precision measurements and how to collaborate with other people outside of the Maryland group. I especially thank Dr. Seidel as well for reading and commenting on the thesis draft.

I would like to thank Dr. Pedro Silva and Dr. Jan Steggemann, for providing invaluable help and guidance in W recoil and DeepMET studies. I am always amazed by their critical insights and deep understandings of physics problems, from which I have learned a lot.

Also thank all my friends at the Maryland group: Dr. Long Wang, Young, Kak, Sara, Yi-Mu, Christos, and Yihui. It is a great experience working together with you in the same group.

And thank my friends at Maryland: Yiming, Peizhi, Fangli, Yu, Shukai, Zishuo, Yipeng, Fan, and friends at CERN: Zhicai, Xunwu, Rongkun, Xiong, Miao, Jiajing, and Xu. Days would be boring without talking and hanging out with you.

Last, I thank my parents for raising me for more than 20 years and providing me the guidance and love that I happily accepted. And I thank my girlfriend Yubo for her company and support and encouragement during my times of depression.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ALICE	A Large Ion Collider Experiment
APD	Avalanche Photodiode
ATLAS	A Toroidal LHC ApparatuS
BSM	Beyond the Standard Model
CDF	Cumulative Distribution Function
CERN	European Organization for Nuclear Research
CMS	Compact Muon Solenoid
CSEV	Conversion Safe Electron Veto
CSC	Cathode Strip Chamber
CTF	Combinatorial Track Finder
DNN	Deep Neural Network
DT	Drift Tube
EB	ECAL Barrel
ECAL	Electromagnetic Calorimeter
EE	ECAL Endcap
EM	Electromagnetic
ES	ECAL Preshower
GSF	Gaussian Sum Filter
HAD	Hadronic
HB	HCAL Barrel
HCAL	Hadronic Calorimeter
HE	HCAL Endcap
HF	HCAL Forward
HLT	High-Level Trigger
HO	HCAL Outer
HPD	Hybrid Photodiode
L1	Level-1
LEP	Large Electron-Positron Collider
LHC	Large Hadron Collider
LHCb	Large Hadron Collider beauty
LO	Leading Order
LV	Leading Vertex
MB	Muon Barrel
MC	Monte Carlo
ME	Muon Endcap
MET	Missing Transverse Momentum
ML	Machine Learning
NLO	Next-to-Leading Order
NN	Neural Network

PDF	Parton Distribution Function
PF	Particle Flow
PMT	Photomultiplier Tube
PS	Proton Synchrotron
PSB	Proton Synchrotron Booster
PSV	Pixel Seed Veto
PU	Pileup
PUPPI	Pileup Per Particle Identification
PV	Primary Vertex
QCD	Quantum Chromodynamics
QED	Quantum Electrodynamics
RF	Radio Frequency
RMS	Root Mean Square
RPC	Resistive Plate Chamber
SC	Supercluster
SiPM	Silicon Photomultiplier
SM	Standard Model
SPS	Super Proton Synchrotron
TIB	Tracker Inner Barrel
TOB	Tracker Outer Barrel
VPT	Vacuum Phototriode
WLS	Wavelength shifting

Chapter 1: Introduction

Particle physics studies the nature of subatomic particles that constitute the matter and radiation in our universe. The field has been growing rapidly since the discovery of the first subatomic particle - the electron in 1897 [1]. Figure 1.1 includes some important discoveries in the history of particle physics. The current dominant theory that provides a classification of these fundamental particles and describes their interactions is called the standard model (SM). The SM has successfully accommodated the existence of the W and Z bosons, gluons, the charm, bottom and top quarks, and predicted their properties with high precision, which agreed with the experimental results. The discovery of the Higgs boson [2, 3] in 2012 at the CERN Large Hadron Collider (LHC) marked the last missing ingredient of the SM.

The SM Lagrangian describes the dynamics of particle physics with only 19 free parameters. Despite the huge success of its predictions, which agree with the experimental results exceptionally well, there are still remaining questions that the SM can not answer, such as the origin and nature of dark matter and dark energy, neutrino masses, and gravitation. Two experimental approaches are undertaken to study particle physics. The first is to measure as precisely as possible the SM parameters and check if they are consistent with the SM predictions (precision mea-

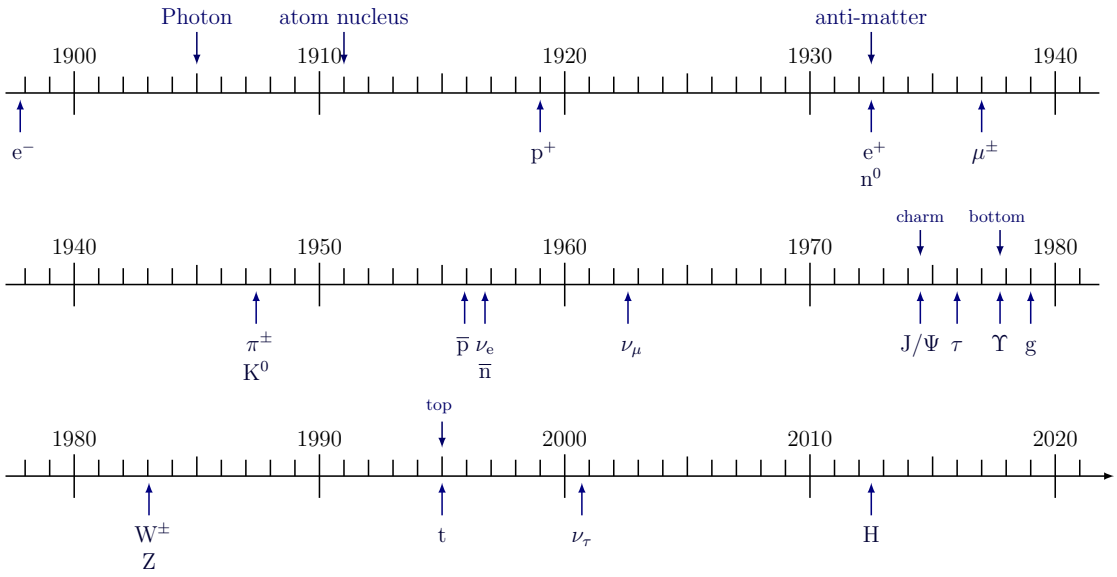


Figure 1.1: Timeline of some important particle discoveries. The figure is adapted from Reference [4] with information from Reference [5].

surements); the second is to look for new particles and new interactions that are not part of the SM (beyond-the-SM searches). The Compact Muon Solenoid (CMS) detector located at the CERN LHC is a multi-purpose detector, that provides the possibility to carry out both approaches simultaneously.

Two proton beams running in opposite directions are accelerated to 6.5 TeV in the LHC accelerator, and brought to collisions at the center of the CMS detector. The properties of the particles produced in these collisions are measured in the detector, and the measurements are used to reconstruct physical objects, such as leptons, photons, and jets. The missing transverse momentum ($p_{\mathrm{T}}^{\mathrm{miss}}$, MET) is one of the key physical objects. It plays a crucial role in interpreting particles that have no interactions within the detector, which is relevant for both precision measurements and beyond-the-SM searches.

Modern deep-learning techniques have been explored within CMS, with great

improvements observed relative to traditional techniques in areas such as heavy flavor jet-tagging algorithms [6, 7], b jet energy measurement [8], and tau lepton identification [9]. However, deep-learning algorithms envisioned only limited application to the $p_{\mathrm{T}}^{\mathrm{miss}}$ estimation.

This dissertation presents a Deep-Neural-Network–based (DNN-based) missing transverse momentum estimator, called "DeepMET". This new estimator operates on each particle individually and assigns a weight w_i and two biases $b_{i,x}, b_{i,y}$ to each particle. The estimator is defined as the negative weighted momentum sum of all particles in an event together with their bias contributions. A 10-20% improvement has been observed compared with the current best $p_{\mathrm{T}}^{\mathrm{miss}}$ estimator.

The dissertation also explores the application of DeepMET in the study of the hadronic recoil against the W boson in $\mathrm{W} \to \ell\nu$ events. W and Z bosons are produced predominantly through quark-antiquark annihilation in pp collisions. Higher-order processes can include radiated gluons or quarks that recoil against the boson and introduce non-zero boson transverse momentum [10]. Figure 1.2 shows an example diagram of such processes. The hadronic recoil serves as an important element in the measurement of the W mass, which is one of the most fundamental parameters of the SM. Measuring the W mass provides a highly stringent check of the validity of the SM, and constrains the parameters of many models beyond the SM (BSM) [11, 12]. Systematic uncertainties from the W transverse momentum on the W mass measurement are studied, in which the DeepMET estimator shows significant improvements relative to other estimators.

The dissertation is organized as follows. Chapter 2 gives a brief overview of the

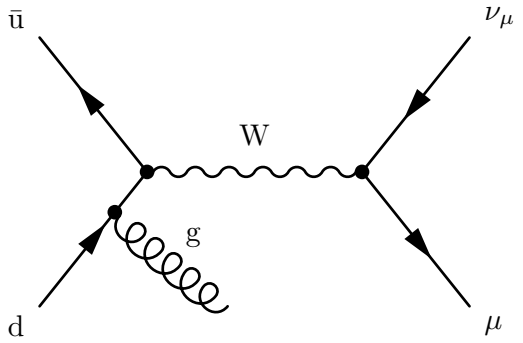


Figure 1.2: An example of a diagram for the production and leptonic decay of the W boson with a radiated gluon in pp collisions.

SM and the global electroweak fit to the SM parameters. Chapters 3 and 4 describe the experimental setup, focusing on the LHC accelerator and the CMS detector, respectively. Chapter 5 is devoted to describing the object and event reconstruction algorithms utilized in the CMS experiment. Chapter 6 presents the new DNN-based $p_{\mathrm{T}}^{\mathrm{miss}}$ estimator DeepMET. Chapter 7 discusses an important application of DeepMET in the measurement of the hadronic recoil against the W boson. Finally, Chapter 8 summarizes the studies advanced in this dissertation, and presents an outlook for future developments.

Chapter 2: Theory

The Standard Model (SM) [13–17] of particles physics is a theoretical framework that describes all known elementary particles and three of four fundamental forces (the electromagnetic force, the weak force, and the strong force, except for gravity) in the universe. It is a renormalizable quantum field theory with local gauge symmetries, represented with the gauge group $\mathrm{SU(3)_C} \times \mathrm{SU(2)_L} \times \mathrm{U(1)_Y}$.

The SM divides all elementary particles into three classes: spin-1/2 fermions, gauge bosons and the Higgs boson. Spin-1/2 fermions include three generations of quarks ((u, d), (c, s), and (t, b)) and three generations of leptons ((e, ν_e), (μ, ν_μ), (τ, ν_τ)). Quarks carry color charge and thus participate in strong interactions. Leptons are colorless and do not experience strong interactions. All quarks and leptons obey Fermi-Dirac statistics and are the fundamental elements that make up matter. Gauge bosons (g, γ, W, and Z) have spin-1, obey Bose-Einstein statistics, and serve as force carriers mediating the three fundamental interactions between particles. The Higgs boson is the only spin-0 scalar boson; it explains how fermions and the W and Z gauge bosons acquire masses. Figure 2.1 summarizes the elementary particles in the SM, together with the electrical charge, mass, and spin of each particle. Figure 2.2 summarizes different kinds of interactions allowed in the SM.

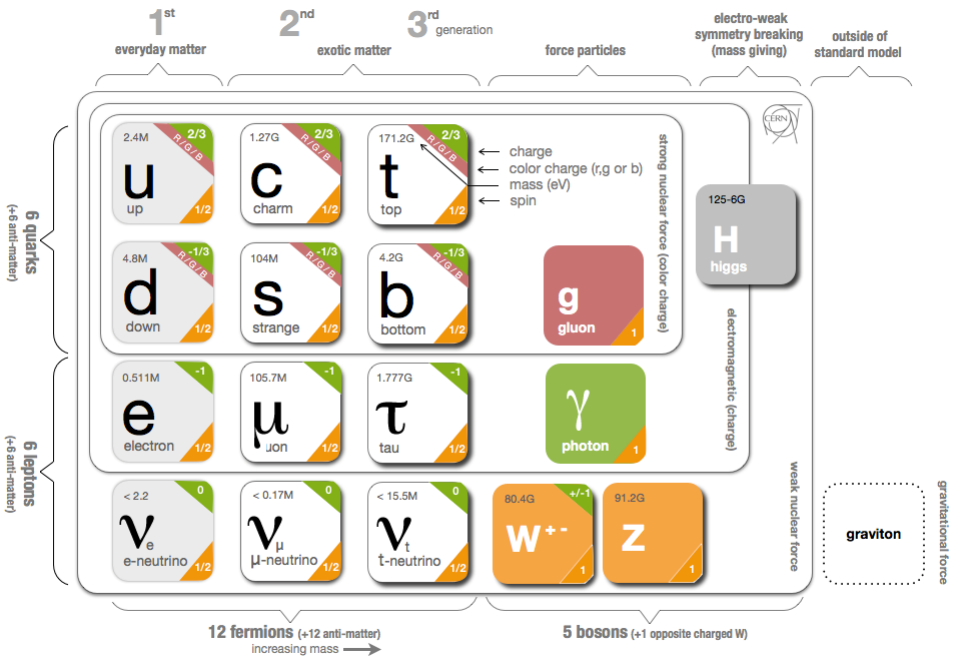


Figure 2.1: Summary of elementary particles in the SM, together with the electrical charge, mass, and spin of each particle [18].

2.1 Quantum Electrodynamics

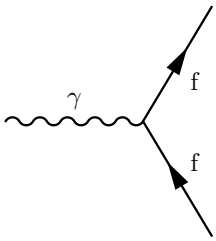
Quantum electrodynamics (QED) is the relativistic quantum field theory of electrodynamics. It is an Abelian gauge theory with the symmetry group U(1).

The Lagrangian of a relativistic spin-1/2 free Dirac fermion can be written as:

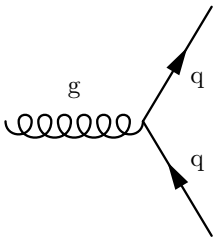
$$\mathcal{L}_{\text{Dirac}} = \bar{\psi}(x)(i\gamma^{\mu}\partial_{\mu} - m)\psi(x)\,, \tag{2.1}$$

where $\psi(x)$ is a Dirac spinor, γ^{μ} is the Dirac matrix, and $\bar{\psi}(x) = \psi^{\dagger}\gamma^{0}$. The Lagrangian is not invariant under a *local* U(1) gauge transformation:

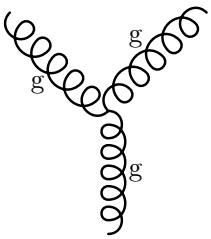
$$\psi'(x) = exp\{iq\theta(x)\}\psi(x)\,, \tag{2.2}$$



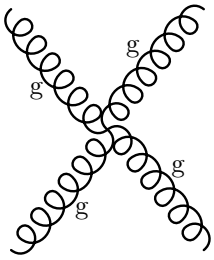
(a) f is any electrically charged fermion



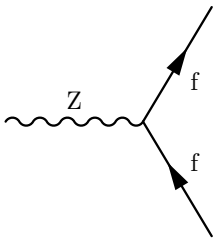
(b) q is any quark



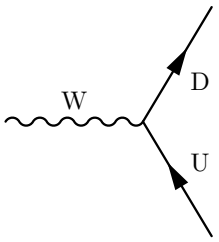
(c) Gluon self interaction



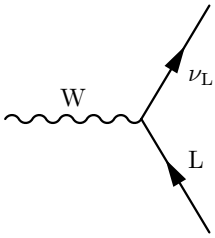
(d) Gluon self interaction



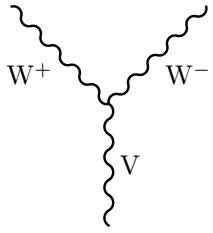
(e) f is any fermion



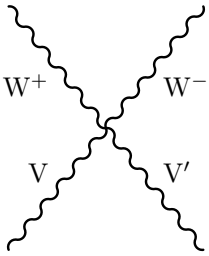
(f) U is any up-type quark; D is any down-type quark



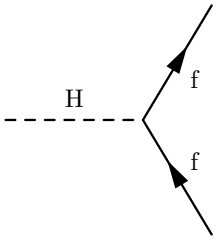
(g) L is a lepton and ν_L is the corresponding neutrino



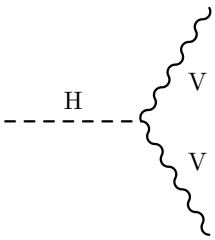
(h) V is a photon or Z boson



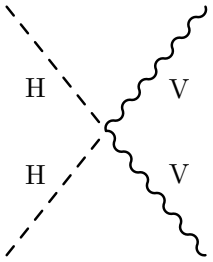
(i) V and V′ are W/Z/γ with charge conservation



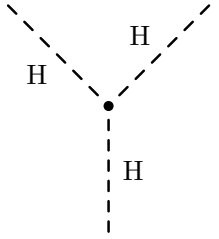
(j) f is a massive fermion



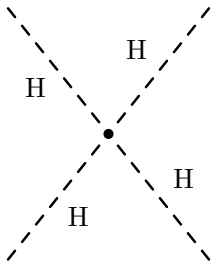
(k) V is a massive gauge boson



(l) V is a massive gauge boson



(m) Higgs self interaction



(n) Higgs self interaction

Figure 2.2: Summary of interactions in the SM.

due to the derivative in Equation 2.1. In order to fix this problem, a new spin-1 field $A_\mu(x)$ is introduced, which transforms as:

$$A'_\mu(x) = A_\mu(x) + \frac{1}{e}\partial_\mu\theta\,. \tag{2.3}$$

The covariant derivative is defined as:

$$D_\mu\psi(x) = [\partial_\mu - ieqA_\mu(x)]\psi(x)\,, \tag{2.4}$$

which transforms like the field itself: $(D_\mu\psi(x))' = exp\{iq\theta\}D_\mu\psi(x)$. Consequently, the new Lagrangian, which is written as:

$$\mathcal{L} = \bar{\psi}(x)(i\gamma^\mu D_\mu - m)\psi(x) = \mathcal{L}_{\text{Dirac}} + eqA_\mu(x)\bar{\psi}(x)\gamma^\mu\psi(x)\,, \tag{2.5}$$

is invariant under local U(1) transformations. The gauge-invariant kinetic term:

$$\mathcal{L}_{\text{Kin}} = -\frac{1}{4}F_{\mu\nu}(x)F^{\mu\nu}(x)\,, \tag{2.6}$$

where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, should be added to the Lagrangian in order to make A_μ a true propagating field. The total Lagrangian is:

$$\begin{aligned}\mathcal{L}_{\text{QED}} &= \mathcal{L}_{\text{Dirac}} + \mathcal{L}_{\text{Kin}} + \mathcal{L}_{\text{int}} \\ &= \bar{\psi}(x)(i\gamma^\mu\partial_\mu - m)\psi(x) - \frac{1}{4}F_{\mu\nu}(x)F^{\mu\nu}(x) + eqA_\mu(x)\bar{\psi}(x)\gamma^\mu\psi(x)\,.\end{aligned} \tag{2.7}$$

The $\mathcal{L}_{\text{int}}$ describes electromagnetic interactions, which are also shown in the second

plot of the first row in Figure 2.2. The fine-structure constant, characterizing the electromagnetic coupling strengths, is defined as $\alpha = e^2/4\pi$.

The gauge invariance constraint forbids a mass term $\frac{1}{2}m^2 A_\mu A^\mu$, making this field massless.

2.2 Quantum Chromodynamics

Quantum chromodynamics is the theory of strong interactions between quarks and gluons. It is a non-Abelian gauge theory, with the $SU(3)_C$ symmetry group. The subscript C represents the color, and there are three types of color charges in total: red, green, and blue.

The free Lagrangian of the quark fields can be written as:

$$\mathcal{L}_0 = \sum_f \bar{q}_f (i\gamma^\mu \partial_\mu - m_f) q_f \,, \tag{2.8}$$

where f represents the quark flavor and $q_f^T = (q_f^1, q_f^2, q_f^3)$ is a vector in color space. The local SU(3) gauge transformation is written as:

$$U = exp\{i\frac{\lambda^a}{2}\theta_a\} \,, \tag{2.9}$$

where $\frac{\lambda^a}{2}$ $(a = 1, 2, ..., 8)$ are the SU(3) generators and satisfy the relation:

$$[\frac{\lambda^a}{2}, \frac{\lambda^b}{2}] = if^{abc}\frac{\lambda^c}{2} \,, \tag{2.10}$$

with f^{abc} denoting the SU(3) structure constants. Similar to the QED case discussed in Section 2.1, the gluon gauge bosons $G_a^\mu(x)$ are introduced to modify the derivatives in order to make the Lagrangian invariant under local SU(3) transformations:

$$D^\mu q_f = [\partial^\mu - i g_s \frac{\lambda^a}{2} G_a^\mu(x)] q_f \,, \tag{2.11}$$

where the factor g_s is the QCD coupling constant. A more often used quantity is $\alpha_{\mathrm{S}} = g_s^2/4\pi$, referred to as "strong coupling constant", analogous to the fine-structure constant α in QED. The corresponding $G_a^\mu(x)$ field strength is described as:

$$G_a^{\mu\nu}(x) = \partial^\mu G_a^\nu - \partial^\nu G_a^\mu + g_s f^{abc} G_b^\mu G_c^\nu \,. \tag{2.12}$$

Finally, the QCD Lagrangian that is invariant under local SU(3) transformations is:

$$\begin{aligned}
\mathcal{L}_{\mathrm{QCD}} =& -\frac{1}{4} G_a^{\mu\nu} G^a_{\mu\nu} + \sum_f \bar{q}_f (i\gamma^\mu D_\mu - m_f) q_f \\
=& -\frac{1}{4} (\partial^\mu G_a^\nu - \partial^\nu G_a^\mu)(\partial_\mu G^a_\nu - \partial_\nu G^a_\mu) + \sum_f \bar{q}_f (i\gamma^\mu \partial_\mu - m_f) q_f \\
&+ g_s G_a^\mu \sum_f \bar{q}_f^\alpha \gamma_\mu (\frac{\lambda^a}{2})_{\alpha\beta} q_f^\beta \\
&- \frac{g_s}{2} f^{abc} (\partial^\mu G_a^\nu - \partial^\nu G_a^\mu) G_\mu^b G_\nu^c - \frac{g_s^2}{4} f^{abc} f_{ade} G_b^\mu G_c^\nu G_\mu^d G_\nu^e \,,
\end{aligned} \tag{2.13}$$

where the first line denotes the kinetic terms for quarks and gluons and the second line denotes the interactions between quarks and gluons. Differently from QED, there are gluon self-interaction terms in the third line because of the QCD's non-

Abelian character. The three interaction terms are shown in Figure 2.2(b), Figure 2.2(c),and Figure 2.2(d), respectively.

Due to gluon self-interactions, the renormalized strong coupling constant α_S has a unique behavior: it is small at high energy (short distance) (*asymptotic freedom*) and large at low energy (large distance). Figure 2.3 shows the measured evolution of α_S as a function of the energy scale Q. When two bare quarks are created and travel apart from each other, the potential energy stored in the gluon field mediating the interactions get larger. Eventually there will be quark-antiquark pairs popping up from the vacuum and forming color-neutral hadrons with the original quarks, leaving no isolated quark or gluon (*color confinement*).

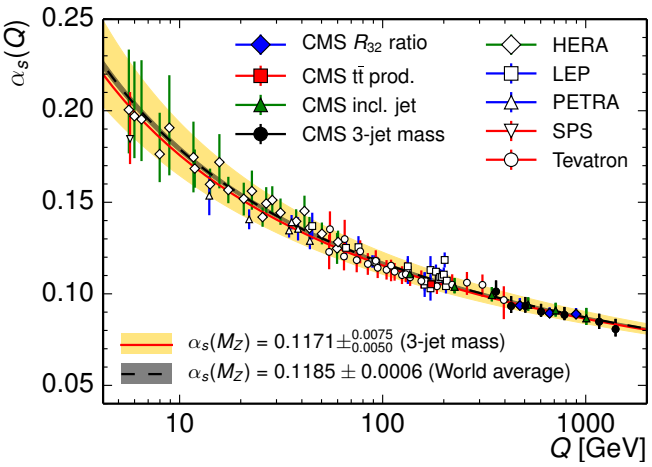


Figure 2.3: Comparison of the $\alpha_S(Q)$ evolution as a function of the energy scale Q [19].

2.3 The Electroweak Theory

The electroweak theory unifies the electromagnetic and the weak forces mediated by the γ, W, and Z gauge bosons. The gauge group is $SU(2)_L \times U(1)_Y$, where

the subscript L refers to left-handed fields and Y refers to hypercharge.

In the SM, quarks and leptons form the left-handed doublet and right-handed singlet. For instance, the first generation of leptons e and $\nu_{\rm e}$ can be described as:

$$L_{\rm e} = \begin{pmatrix} \nu_{{\rm e},L} \\ \\ {\rm e}_L \end{pmatrix} \quad \text{and} \quad {\rm e}_R, \tag{2.14}$$

with the left and right handed states defined as:

$${\rm e}_L = \frac{1-\gamma^5}{2}{\rm e}, \qquad \nu_{{\rm e},L} = \frac{1-\gamma^5}{2}\nu_{\rm e}, \qquad {\rm e}_R = \frac{1+\gamma^5}{2}{\rm e}\,, \tag{2.15}$$

where γ^5 is a 4×4 matrix $\left(\begin{smallmatrix} 0 & I_{2\times 2} \\ I_{2\times 2} & 0 \end{smallmatrix}\right)$. Similar ideas apply to the quark sector: the first generation of quarks u and d can be written as: $Q_1 = \begin{pmatrix} {\rm u}_L \\ \\ {\rm d}_L \end{pmatrix}$, ${\rm u}_R$, and ${\rm d}_R$.

Analogous to the QED and QCD cases, in order to keep the Lagrangian invariant under the gauge transformation of $\mathrm{SU(2)_L \times U(1)_Y}$, the derivatives should be:

$$\begin{aligned} D_{\mu,L} &= \partial_\mu - ig\frac{\sigma_i}{2}W^i_\mu(x) - ig' y_1 B_\mu(x)\,, \\ D_{\mu,R} &= \partial_\mu - ig' y_2 B_\mu(x)\,, \end{aligned} \tag{2.16}$$

where g and g' represent the gauge couplings for the SU(2) and U(1) group, respectively; $\sigma_i/2$ is the generators of the SU(2) group; y_i is the hypercharge. With this the Lagrangian for the fermion field is invariant under gauge transformations.

The Lagrangian for the gauge field is:

$$\mathcal{L}_{\text{kin}} = -\frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}W^i_{\mu\nu}W_i^{\mu\nu}\,, \tag{2.17}$$

with the field strengths:

$$\begin{aligned} B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu\,, \\ W^i_{\mu\nu} &= \partial_\mu W^i_\nu - \partial_\nu W^i_\mu + g\epsilon^{ijk}W^j_\mu W^k_\nu\,. \end{aligned} \tag{2.18}$$

The total Lagrangian is the sum of the fermion fields and the gauge fields:

$$\begin{aligned} \mathcal{L}_{\text{EW}} &= \mathcal{L}_f + \mathcal{L}_{\text{kin}} \\ &= \bar{L}_{\text{e}} i\gamma^\mu D_\mu L_{\text{e}} + \bar{\text{e}}_R i\gamma^\mu D_\mu \text{e}_R + ... - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}W^i_{\mu\nu}W_i^{\mu\nu}\,. \end{aligned} \tag{2.19}$$

The covariant derivatives in Equation 2.16 introduce the interactions between the fermions and the gauge bosons, which can be split into two categories: the off-diagonal charged-current interactions, denoted by the boson fields:

$$\begin{aligned} W_\mu &= (W^1_\mu + iW^2_\mu)/\sqrt{2}\,, \\ W^\dagger_\mu &= (W^1_\mu - iW^2_\mu)/\sqrt{2}\,, \end{aligned} \tag{2.20}$$

and the diagonal neutral-current interactions, denoted by W^3_μ and B_μ. The W^3_μ and

B_μ are combined to form the γ (A_μ) and Z bosons:

$$\begin{aligned} A_\mu &= B_\mu \cos\theta_{\mathrm{W}} + W_\mu^3 \sin\theta_{\mathrm{W}} \,, \\ Z_\mu &= -B_\mu \cos\theta_{\mathrm{W}} + W_\mu^3 \sin\theta_{\mathrm{W}} \,, \end{aligned} \tag{2.21}$$

where θ_{W} is the Weinberg weak-mixing angle, with:

$$\cos\theta_{\mathrm{W}} = \frac{g}{\sqrt{g^2 + g'^2}} \,. \tag{2.22}$$

The electric charge should satisfy:

$$e = g \sin\theta_{\mathrm{W}} = g' \cos\theta_{\mathrm{W}} \,, \tag{2.23}$$

in order to have the A_μ term consistent with QED. The covariant derivatives also generate interactions between different gauge bosons, shown in Figure 2.2(h) and Figure 2.2(i).

2.4 The Higgs Mechanism

It appears that a theory containing only fermions and gauge bosons would need them to be massless, for the gauge symmetries to be preserved. This is contradicted by experimental observations. The Higgs mechanism was introduced [20–24] to solve this problem and explain how fermions and gauge bosons acquire their respective masses.

A doublet of complex scalar fields is introduced in the Higgs mechanism, writ-

ten as:

$$\phi(x) = \begin{pmatrix} \phi^{\dagger}(x) \\ \phi^{0}(x) \end{pmatrix} , \tag{2.24}$$

and the potential for this scalar field (the *Higgs potential*):

$$V(\phi) = \mu^2 \phi^{\dagger}\phi + \lambda(\phi^{\dagger}\phi)^2 , \tag{2.25}$$

in which λ should be larger than 0 for the potential to be bound. For $\mu^2 > 0$, the field has a trivial minimum potential at $\phi = 0$. However, for $\mu^2 < 0$, the field has an infinite set of degenerate states with minimum potential, described by $|\phi|^2 \equiv \nu^2/2 = -\mu^2/2\lambda$ (*vacuum expectation value*). Figure 2.4 shows an illustration of the Higgs potential. Without loss of generality, the ground vacuum state can be chosen as:

$$\langle\phi\rangle_0 = \begin{pmatrix} 0 \\ \nu/\sqrt{2} \end{pmatrix} . \tag{2.26}$$

Once a particular ground state is chosen, the SU(2) × U(1) symmetry gets *spontaneously broken*, while a U(1) gauge symmetry is preserved (which describes electromagnetism).

The field can be rewritten around this ground vacuum:

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \nu + H(x) \end{pmatrix} . \tag{2.27}$$

Together with the covariant derivative $D^{\mu}\phi = [\partial^{\mu} - ig\frac{\sigma^i}{2}W_i^{\mu} - ig'\frac{1}{2}B^{\mu}]\phi$, the kinetic

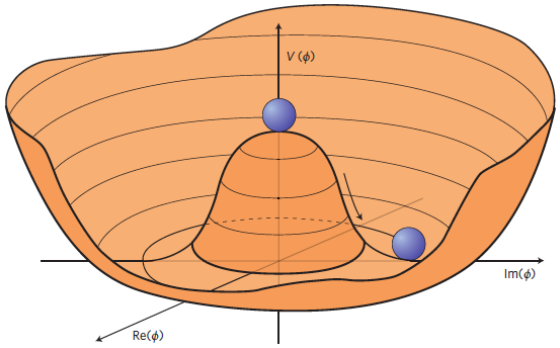


Figure 2.4: Illustration of the Higgs potential, with an infinite set of degenerate states with minimum potential [25].

term of the scalar field Lagrangian is

$$(D_\mu\phi)^\dagger D^\mu\phi = \frac{1}{2}\partial_\mu H\partial^\mu H + (\nu+H)^2\{\frac{g^2}{4}W_\mu^\dagger W^\mu + \frac{g^2}{8\cos^2\theta_\mathrm{W}}Z_\mu Z^\mu\}\,, \tag{2.28}$$

and the total Lagrangian is

$$\begin{aligned}\mathcal{L}_S =&(D_\mu\phi)^\dagger D^\mu\phi - V(\phi)\\ =&+\frac{1}{2}\partial_\mu \mathrm{H}\partial^\mu \mathrm{H} - \frac{\mu^4}{4\lambda}\\ &+\frac{g^2\nu^2}{4}\mathrm{W}_\mu^\dagger \mathrm{W}^\mu\{1+\frac{2}{\nu}\mathrm{H}+\frac{\mathrm{H}^2}{\nu^2}\}+\frac{g^2\nu^2}{8\cos^2\theta_\mathrm{W}}\mathrm{Z}_\mu\mathrm{Z}^\mu\{1+\frac{2}{\nu}\mathrm{H}+\frac{\mathrm{H}^2}{\nu^2}\}\\ &-\mu^2\mathrm{H}^2+\lambda\nu\mathrm{H}^3+\frac{\lambda}{4}\mathrm{H}^4\,,\end{aligned} \tag{2.29}$$

where the quadratic term for the Higgs boson, and the $\mathrm{W}^\pm$ and the Z gauge bosons are created, with the boson masses:

$$m_\mathrm{H} = \sqrt{-2\mu^2} = \sqrt{2\lambda\nu^2}, \qquad m_\mathrm{W} = \frac{g\nu}{2} \quad \text{and} \quad m_\mathrm{Z} = \frac{g\nu}{2\cos\theta_\mathrm{W}} \tag{2.30}$$

Photons remain massless since there is no mass term created for A_μ. Equation 2.29

also includes the interactions of HWW, HHWW, HZZ, HHZZ, and the Higgs boson self interactions H^3 and H^4. The HWW coupling strength is $2m_W^2/\nu$ and the HZZ coupling strength is m_Z^2/ν, both proportional to the square of gauge boson masses.

Fermion masses can also be acquired from the interactions with the Higgs field via Yukawa couplings. It can be written as:

$$\mathcal{L}_Y = -(1 + \frac{H}{\nu})m_f \bar{f} f \tag{2.31}$$

with the coupling strength m_f/ν between the Higgs boson and the fermions, proportional to the fermion mass.

The interactions involving the Higgs boson are shown in Figure 2.2(j) to Figure 2.2(n).

2.5 The Global Electroweak Fit

The SM parameters constrain one another with the above calculations. There are only 7(+1) free parameters in the electroweak sector for one fermion generation. The choice of these free parameters is not unique, and one example can be the two gauge couplings g and g', the two parameters μ^2 and λ in the Higgs potential, and masses of fermions m_u, m_d, m_e (and m_{ν_e}). The weak mixing angle θ_W, the W, Z, and the Higgs boson masses, the Yukawa coupling strengths, and other parameters can then be determined from calculations and compared with (in)direct experimental measurements.

The global electroweak fit exploits this method. It combines measurements

of different SM parameters in various processes into one single fit, to obtain the precise evaluation of the SM parameters and probe the stringent test of the SM. It successfully predicted the top quark [26] and the Higgs boson [11] mass windows before their discoveries. The inputs to the fits include the Higgs boson mass m_{H}, the W boson mass m_{W} and width Γ_{W}, the Z boson mass m_{Z} and width Γ_{Z}, the effective weak-mixing angle $\sin^2\theta^f_{\mathrm{eff}}$, the top quark mass m_{t}, etc. More detailed information and the latest fit results are provided in [12].

The W boson mass m_{W} is a key ingredient among the input variables. It follows the relation [27]:

$$m_{\mathrm{W}}^2\left(1-\frac{m_{\mathrm{W}}^2}{m_{\mathrm{Z}}^2}\right)=\frac{\pi\alpha}{\sqrt{2}G_\mu}(1+\Delta r)\,, \tag{2.32}$$

where $G_\mu = 1.17\times10^{-5}\,\mathrm{GeV}^2$ is the Fermi constant; Δr represents the higher-order corrections, and is sensitive to the top quark and the Higgs boson masses, as well as new particles and interactions from theories beyond the SM. Figure 2.5 provides the comparison of the constraints on m_{H} from indirect fit results and the direct measurements, where the m_{W} provides one of the strongest constraints. Figure 2.6 is an example of the SM consistent test between the indirect determinations of m_{W}, m_{t} and the direct experiment measurement results.

Figure 2.7 shows the $\Delta\chi^2$ scan as a function of m_{W}. The uncertainty on m_{W} from indirect fit is about 7 MeV, while the combined experimental uncertainty is about 13 MeV, larger than the SM theoretical prediction.

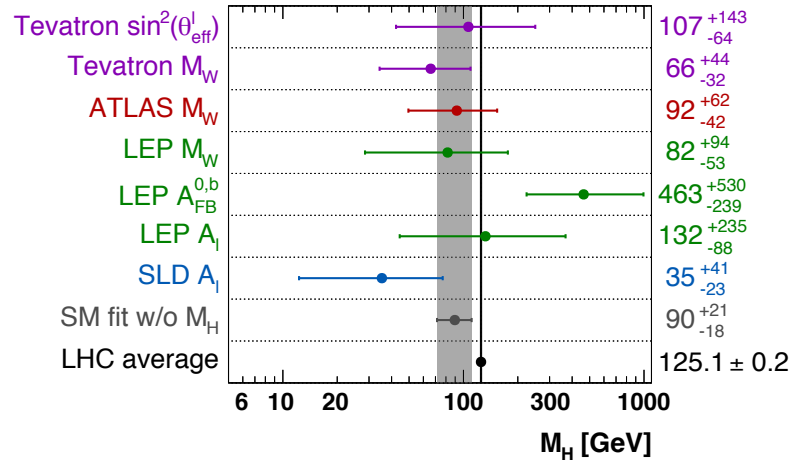


Figure 2.5: Comparison of the constraints on m_{H} obtained indirectly from individual observables with the fit result and the direct LHC measurement [12].

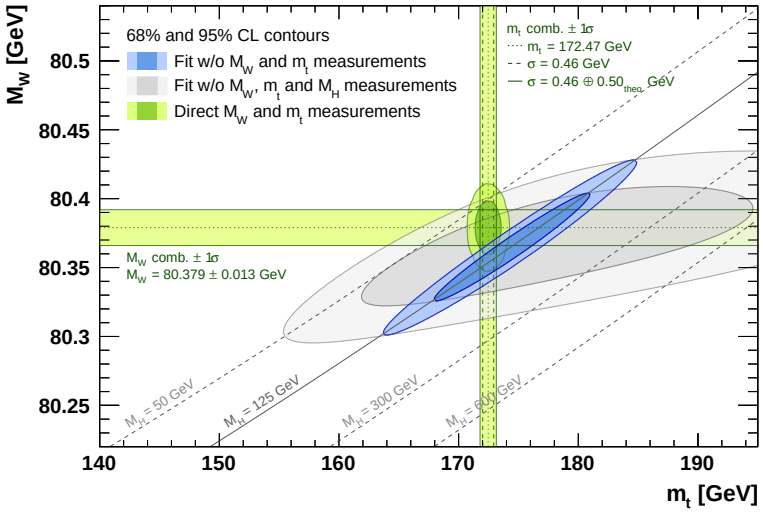


Figure 2.6: Contours at 68% and 95% CL obtained from scans of m_{W} *vs.* m_{t} for the fit including (blue) and excluding (gray) the m_{H} measurement, as well as the direct measurements of m_{W} and m_{q} [12].

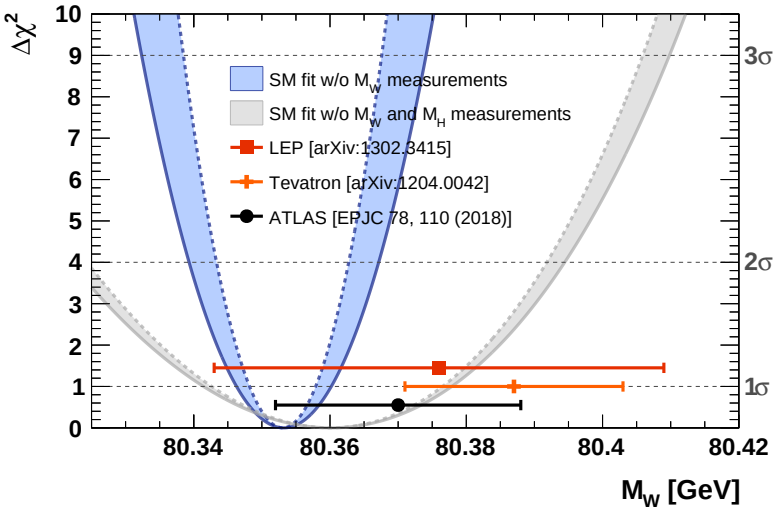


Figure 2.7: Scan of the $\Delta\chi^2$ as a function of m_{W} [12]. The current uncertainty from the SM fit is about 7 MeV, smaller than the experimental measurements.

Chapter 3: The Large Hadron Collider

The Large Hadron Collider (LHC) [28] is the world's largest circular particle accelerator that accelerates and collides beams of protons or heavy ions at the highest controlled energy and largest instantaneous luminosity. It is installed in the 26.7 km tunnel located 45-170 m beneath the surface in the border between France and Switzerland. The tunnel was completed in 1988 for the CERN Large Electron-Position Collider (LEP) machine [29].

The LHC consists of eight independent sectors (octants), each having one straight section in its center and two half arcs at the two sides. Each straight section is approximately 528 m long and can serve as an experimental or utility insertion. Among these eight possible interaction regions, four have beam crossings; an independent experiment is installed in each of them, as shown in Figure 3.1. Two high-luminosity multi-purpose detectors, ATLAS [30] and CMS [31], are located at Point 1 and Point 5, directly across the ring from one another to ensure approximately equivalent delivered luminosity. ALICE [32], a dedicated detector for heavy ion collisions, operates at Point 2 and LHCb [33], a dedicated b physics detector, operates at Point 8. The injection systems for Beam 1 and Beam 2 are also located in Point 2 and Point 8, respectively. The remaining four straight sections do not

have beam crossings and serve for other purposes: two collimated systems, one in Point 3 and the other in Point 7; two radio frequency (RF) systems in Point 4, one for each LHC beam; the beam dump systems in Point 6, which allow for each beam to be aborted independently.

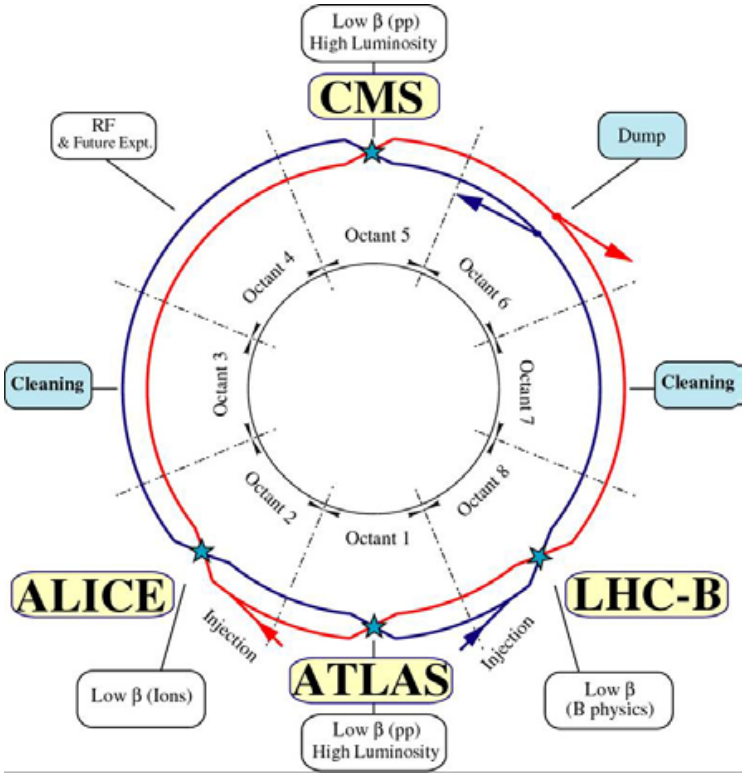


Figure 3.1: Schematic layout of the LHC [28].

The LHC is designed to collide proton beams with a center-of-mass energy $\sqrt{s}$ of 14 TeV and luminosity up to $10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$. To accelerate protons to such a high energy, a chain of smaller accelerators is designed, as illustrated in Figure 3.2. Protons are produced in a duoplasmatron source and accelerated to 50 MeV in the Linac2 linear accelerator. They are then injected into the Proton Synchrotron Booster (PSB) and accelerated to 1.4 GeV. The PSB has four independent rings and operates on the first harmonic of the natural wavelength of the RF system (harmonic $h = 1$).

When protons are transferred from the PSB to the Proton Synchrotron (PS) for further acceleration, the LHC two-batch filling scheme is used, in which six bunches from two batches (split by $4 + 2$ or $3 + 3$) of the PSB are transferred to the PS on harmonic $h = 7$, leaving one bunch in the PS empty. The filling process is shown in Figure 3.3. In the PS, each of the six bunches is split into three bunches, accelerated to 25 GeV on harmonic $h = 21$, and then further split into four bunches. Therefore, the six original bunches injected from the PSB have been split into 72 bunches in the PS, on harmonic $h = 84$ with 12 consecutive empty buckets. The empty buckets provide a gap for the rise-time of the ejection kicker. In the next step, the bunch train is injected into the Super Proton Synchrotron (SPS) and accelerated to 450 GeV. Finally, the proton bunches are injected from the SPS into the main LHC ring through the injection system at Point 2 and Point 8.

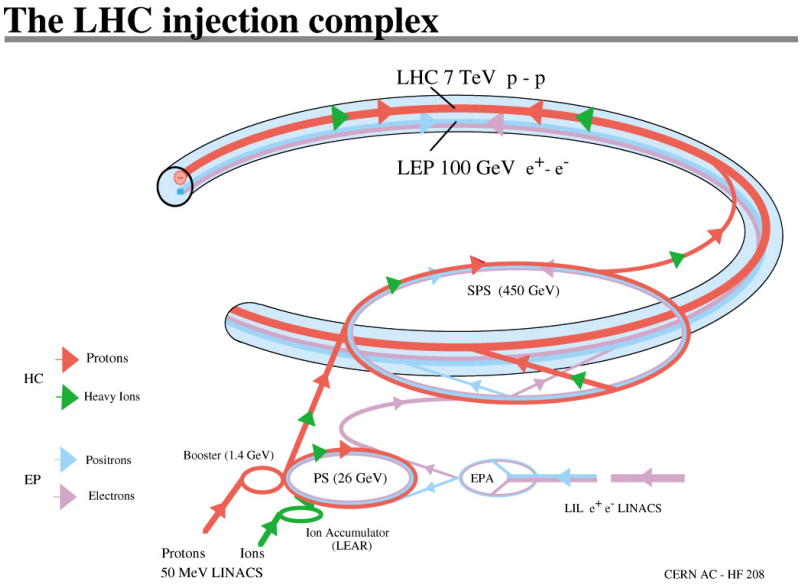


Figure 3.2: LHC injection complex [34].

The harmonic $h = 84$, always referred to as "harmonic 84", originates from

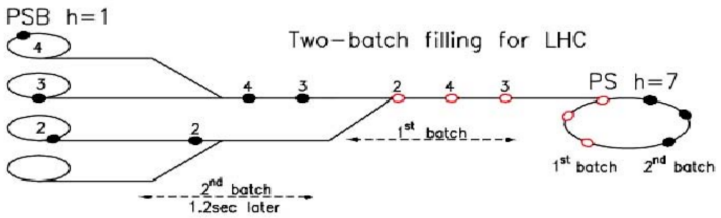


Figure 3.3: The LHC two batch filling scheme for the PSB to the PS transfer. [34].

the requirement of the LHC's 25 ns bunch spacing (40 MHz). The length of the LHC is 26659 m, which translates into 88924 ns orbit time for protons. With 25 ns bunch spacing (24.95 ns to be more precise), the maximum number of bunches in the LHC is $88924/24.95 \cong 3564$. The circumference of the SPS is 11 times that of the PS, and the LHC is 27/7 times that of the SPS. So the 3564 bunches in the LHC are essentially $3564/(27/7 \times 11) = 84$ in the PS. Therefore, the harmonic $h = 84$ is needed in the PS, in order to fulfill the 25 ns bunch spacing in the LHC. Figure 3.4 shows the real proton bunches in the PS, SPS and one LHC ring. 2808 proton bunches are distributed in the 3564 time slots, with 756 missing bunches intended for the rise-time of injections and dumps.

The LHC accelerator is installed in the existing LEP tunnel. The internal diameter is only 3.7 m in the arched sections between each sector , making it difficult to install two completely separate proton rings. To overcome this problem, the LHC adopted the twin-bore magnet design [35], where the two beams are placed in a single mechanical structure. Figure 3.5 shows the cross section of the LHC dipole magnets. The magnets are NbTi superconductors cooled to 1.9 K by superfluid helium. With a total number of 1232 dipole magnets, the maximum magnetic field in the LHC is 8.33 T, allowing for proton energies up to 7 TeV.

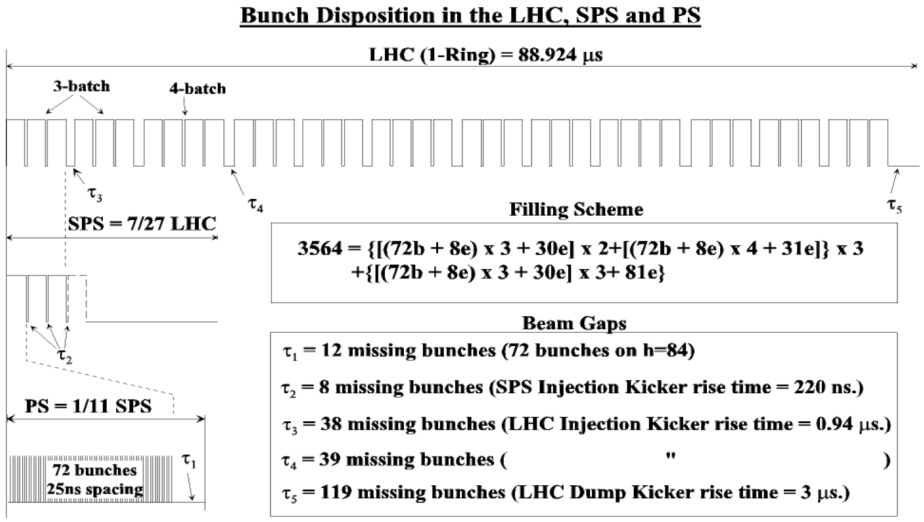


Figure 3.4: Proton bunches in the PS, SPS and one LHC ring [34].

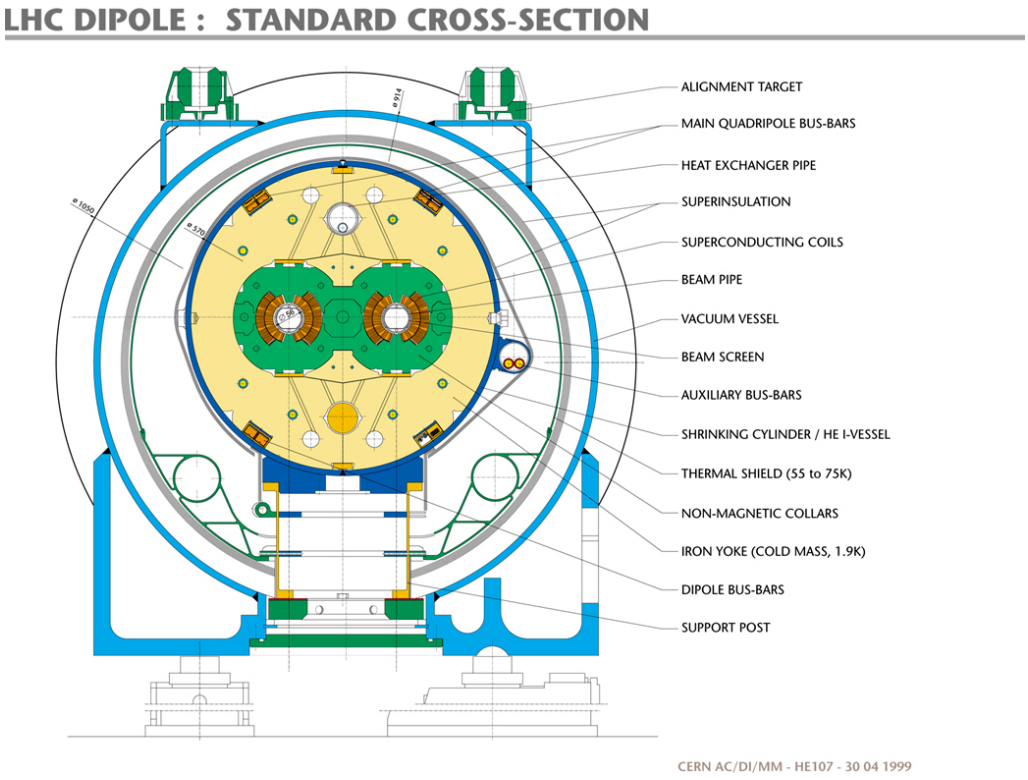


Figure 3.5: Standard cross section of the LHC dipole magnet [36].

The number of events per second generated in the LHC collisions is given by:

$$N_{\text{events}} = L\sigma_{\text{events}}\,, \tag{3.1}$$

where σ_{events} is the cross section for these events and L is the instantaneous luminosity. The instantaneous luminosity depends only on the beam parameters and can be written, for a Gaussian beam distribution, as:

$$L = \frac{N_b^2 n_b f_{\text{rev}} \gamma_r}{4\pi\epsilon_n \beta*} F\,, \tag{3.2}$$

where N_b is the number of particles per bunch; n_b is the number of bunches per beam; f_{rev} is the revolution frequency; γ_r is the relativistic gamma factor; ϵ_n is the normalized transverse beam emittance; $\beta*$ is the beta function at the collision point; F is the geometric luminosity reduction factor due to the crossing angle at the interaction point, and is calculated as:

$$F = (1 + (\frac{\theta_c \sigma_z}{2\sigma*})^2)^{-\frac{1}{2}}\,, \tag{3.3}$$

where θ_c is the full crossing angle between the two beams; σ_z and σ_* are the root mean square (RMS) bunch sizes in the longitudinal and transverse direction, respectively. Table 3.1 lists part of the LHC designed parameters and the parameters during Run 2 (2015-2018) data taking. Figure 3.6 shows the integrated and peak luminosity delivered by the LHC to CMS during stable beams for pp collisions at Run 2. The integrated luminosity delivered by the LHC during Run 2 is 162.85 fb^{-1},

Table 3.1: Typical proton running conditions in the LHC in Run 2 (2015-2018), together with the design parameters. Numbers are mostly taken from Reference [37] and [38].

Parameter	Design	2015	2016	2017	2018
Beam energy [TeV]	7.0	6.5	6.5	6.5	6.5
Protons per bunch N_b [10^{11}]	1.15	1.0-1.25	1.0-1.25	1.0-1.25	1.0-1.25
Number of bunches n_b	2808	2244	2220	2556/1868	2556
Bunch spacing [ns]	25	25	25	25	25
Revolution frequency f_{rev} [Hz]	11245	11245	11245	11245	11245
Relativistic gamma factor γ_{rev}	7461	6928	6928	6928	6928
Transverse emittance ϵ_n [μm]	3.75	3.5	2.2	2.2	1.9
$\beta*$ [m]	0.55	0.8	0.4	0.4-0.3	0.3-0.25
Half crossing angle $\theta_c/2$ [μrad]	143	145	185/140	150-120	160-130
Peak luminosity [$10^{34}\,\text{cm}^{-2}\text{s}^{-1}$]	1.0	0.5	1.4	2.1	2.1

and the integrated luminosity CMS recorded is $150.26\,\text{fb}^{-1}$.

The total inelastic pp cross section at $\sqrt{s} = 13\,\text{TeV}$ is about 80 mb [40, 41]. With a peak instantaneous luminosity of 2.1×10^{34} $\text{cm}^{-2}\text{s}^{-1}$ in 2018, the number of interactions per second was about 1.6×10^9. The LHC proton bunch spacing is 25 ns, which translates into a 40 MHz bunch crossing frequency. The actual frequency is smaller due to the unfilled bunches described previously. In 2018, the frequency was about: $f_{\text{rev}} \times n_b = 11245 \times 2556 = 28.7\,\text{MHz}$. Therefore, the number of pp interactions per bunch crossing was around $1.6 \times 10^9/28.7 = 56$ in 2018. Figure 3.7 shows the mean number of pp interactions per bunch crossing during Run 2. The mean number of interactions was smaller since the average luminosity was smaller. With the trigger selections described later in Section 4.5, there is usually one pp interaction that is related to the process of potential physics interest (*hard scattering*), while other pp interactions from the same bunch crossing (*pileup*) are usually soft and affect the reconstruction of different physics objects.

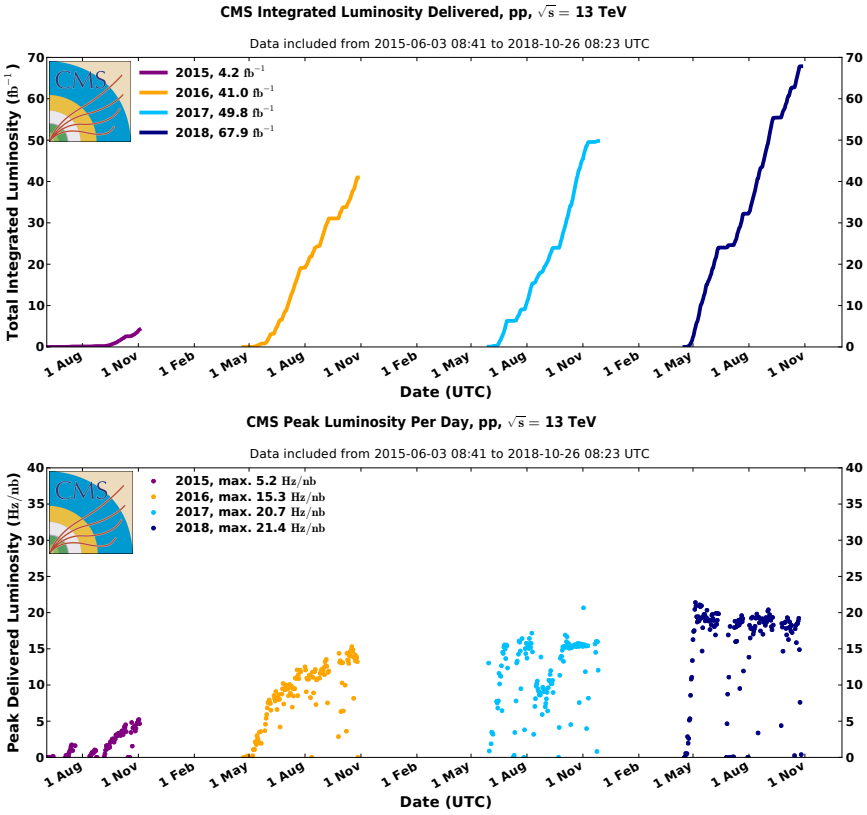


Figure 3.6: Integrated (top) and peak (bottom) luminosity delivered by the LHC during Run 2 pp collisions at $\sqrt{s} = 13\,\text{TeV}$ [39].

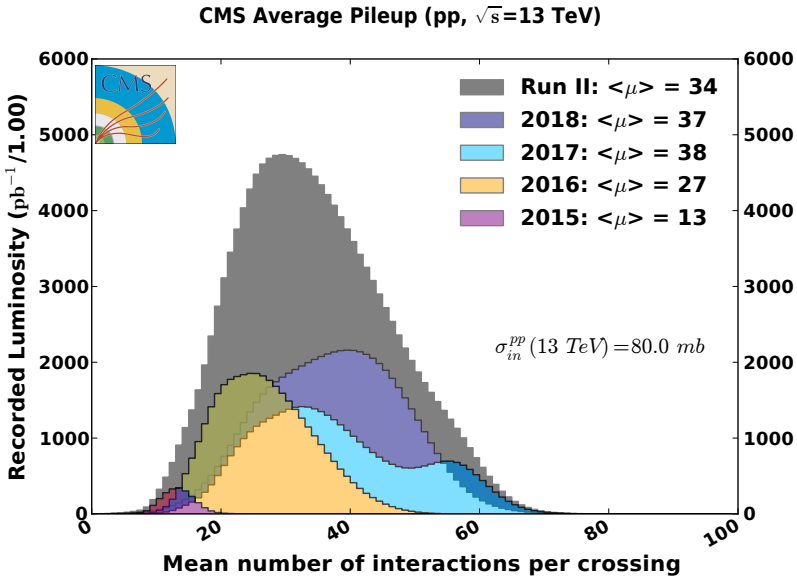


Figure 3.7: Mean number of pp interactions per bunch crossing during Run 2 [39].

Chapter 4: The Compact Muon Solenoid Detector

The main goals of the LHC physics programme include the characterization of the newly discovered Higgs boson candidate, the precise measurements of the SM parameters, and the search for new physics processes beyond the SM. In order to achieve these goals, several requirements need to be satisfied: good identification, and good energy and momentum resolution of muons, electrons and photons, and good energy resolutions of jets and missing transverse momentum. The Compact Muon Solenoid (CMS) detector is well designed to meet these requirements.

The CMS detector is a high-luminosity general purpose detector built 100 m underground at Point 5 of the LHC. The detector has a cylindrical shape, with a radius of 7.5 m, a length of 28.7 m, and a total weight of 14000 tons. It has a long barrel section, and two endcaps on the two ends. Figure 4.1 shows an overview of the CMS detector. The center feature is a 12.5 m long superconducting solenoid with an inner radius of 3 m, which provides a 3.8 T uniform axial magnetic field. This strong magnetic field is crucial for the precise measurement of charged particle momenta. The strength and direction of the magnetic field are shown in Figure 4.2. A highly segmented silicon tracker, a fine-grained crystal electromagnetic calorimeter, and a hermetic brass/scintillator hadron calorimeter are inside the solenoid. The muon

spectrometer embedded in the iron return yokes is outside the solenoid. Figure 4.3 shows a longitudinal view of one quadrant of the CMS detector, with all subsystems and their distances in r-z space from the center plotted.

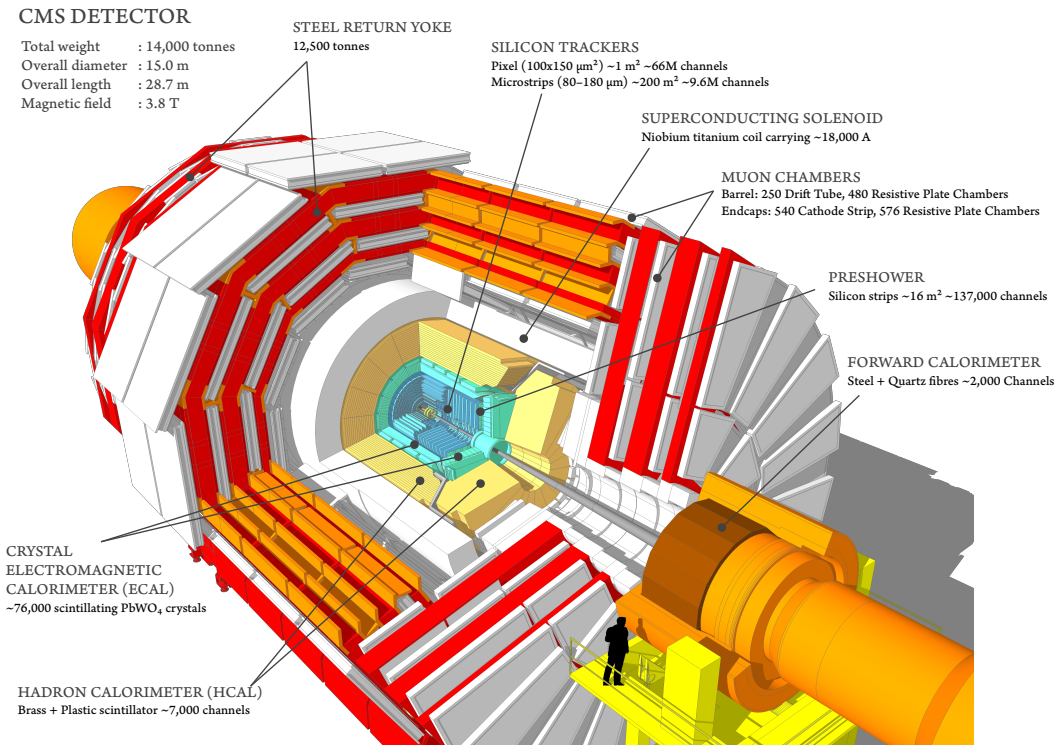


Figure 4.1: The cutaway diagram of the CMS detector [42].

CMS uses a right-handed coordinate system. The origin is located at the nominal interaction point in the center of the CMS detector. The z axis is set along the LHC beam pipe with the $+z$ direction towards the Jura mountain. The x-y plane is perpendicular to the beam pipe, with $+x$ pointing radically toward the center of the LHC ring and $+y$ pointing vertically upward. In the x-y transverse plane, the azimuthal angle ϕ is measured counterclockwise from the positive x axis and the radial coordinate is represented with r. The polar angle θ is measured from the positive z axis. Compared with θ, the pseudorapidity η, defined as $\eta = -\ln\tan(\theta/2)$

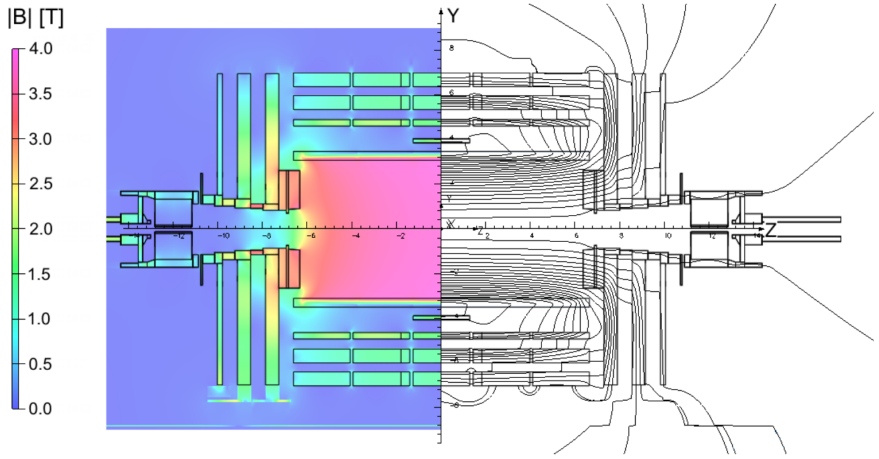


Figure 4.2: Value of $|B|$ (left) and field lines (right) on a longitudinal section of the CMS detector [43].

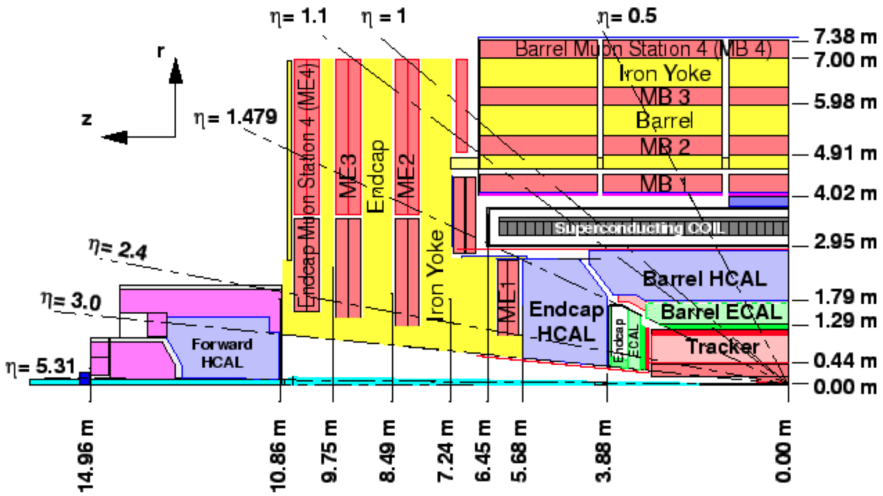


Figure 4.3: Longitudinal view of one quadrant of the CMS [44].

is more regular used. Figure 4.4 illustrates the coordinate system.

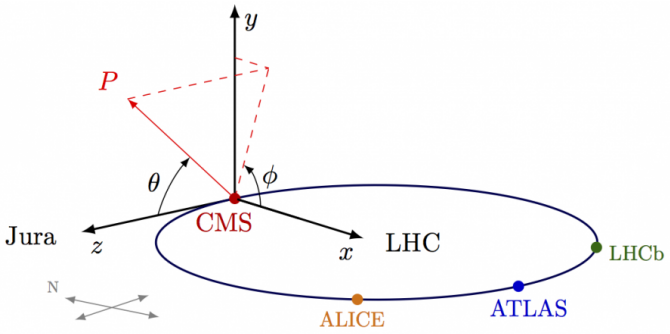


Figure 4.4: Illustration of the CMS coordinate system [45].

The events recorded by the CMS detector are organized following the order of "run, luminosity section, event". One *run* comprises a period with a consistent configuration of the data taking, and can last several hours. One run is divided into a set of *luminosity sections*. One luminosity section is defined to be 2^{18} orbit times (about $2^{18} \times 88924\,\text{ns} = 23.3\,\text{s}$), during which the detector conditions are assumed to be consistent, and the instantaneous luminosity is approximately constant. One luminosity section consists of many *events*, with each event representing one bunch crossing.

The CMS subdetectors will be described in detail in the following sections of this chapter.

4.1 Tracker

The CMS tracking system [46, 47] is the innermost detector. It is designed to carry out precision measurements of positions and momenta of charged particles produced from pp collisions. These measurements are then used for primary and

secondary vertex reconstructions [48]. The tracker has a cylindrical shape of 5.8 m in length and 2.5 m in diameter, with its axis closely aligned to the LHC beam line, and η coverage up to 2.5. Two subsystems make up the tracker: a small inner silicon pixel detector and a large outer strip tracker. Figure 4.5 shows the schematic cross section through the CMS tracker in the r-z plane.

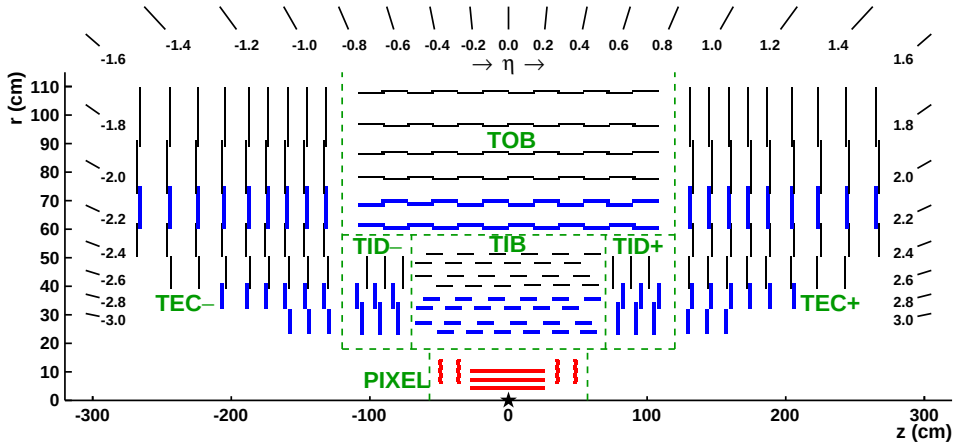


Figure 4.5: Schematic cross section through the CMS tracker in the r-z plane [48]. The tracker is symmetric with respect to the horizontal line $r = 0$, so only the top half is shown here. The red lines represent the pixel tracker; the blue and black lines represent the strip tracker, where the blue is for layers with double-sided strip modules and black is for the ones with single-sided modules. The pixel detector shown in this plot is the one before the Phase-1 pixel upgrade.

The pixel tracker is the closest portion to the beam pipe. Before 2017, it consisted of three barrel layers (BPIX) and two endcap disks (FPIX) in each side. The three BPIX layers were located at radii 4.4, 7.3 and 10.2 cm extending from $z = -26.5$ cm to $z = +26.5$ cm. The two FPIX disks were located at $z = \pm 34.5$ cm and ± 46.5 cm, and covered a region between approximately 6 cm and 15 cm in radius from the beam pipe. In order to meet the increased requirements imposed by high luminosity and pileup, the pixel tracker was upgraded during the technical stop at the beginning of 2017 [49]. The new pixel detector has four barrel layers (located

at radii 3.0, 6.8, 10.9 and 16.0 cm) and three endcap disks at each side (located at $z = \pm 29.1$ cm, ± 39.6 cm, and ± 51.6 cm). Figure 4.6 compares the geometry of the pixel detector before (dashed yellow) and after (solid green) the upgrade. The pixel size is unchanged: $100 \times 150\,\mu\text{m}^2$. The hit position resolution in the pixel tracker is about 10 μm in the transverse coordinate and 20-40 μm in the longitudinal coordinate.

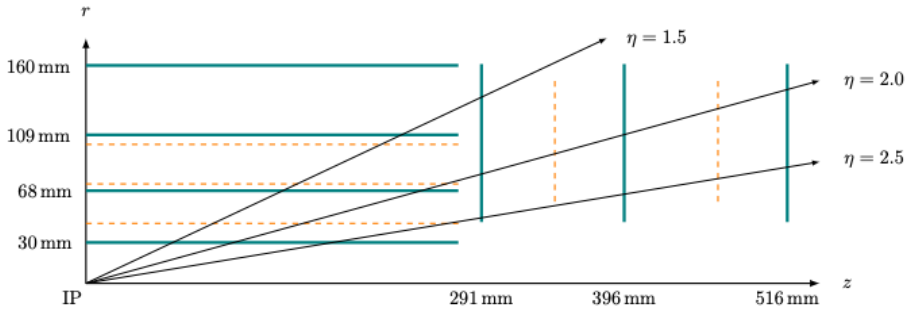


Figure 4.6: Comparison of the pixel detector geometry before and after the Phase-1 upgrade [50]. The dashed yellow is before the upgrade and the solid green is after the upgrade. One quadrant of the detector is shown since it is symmetric along $r = 0$ and $z = 0$.

Surrounding the pixel detector is the strip tracker, occupying the radius between 20 and 116 cm. It consists of four subsystems: tracker inner barrel (TIB) and tracker outer barrel (TOB) with cylindrical shape in the barrel, and tracker inner disk (TID) and tracker endcap (TEC) in the two endcaps. The location and number of layers information is summarized in Table 4.1. The single hit resolution in TIB (TOB) is about 13-38 μm (18-47 μm).

The material budget of the tracking system is optimized in order to minimize the energy loss and multiple scatterings. The estimated total amount of material was 0.4 radiation lengths in the central region and up to 1.8 radiation lengths in the

Table 4.1: Summary of the number of layers and the position information of the four strip tracker subsystem[48].

Subsystem	Layers	Location
TIB	4 cylindrical	$20 < r < 55\,\text{cm}$
TOB	6 cylindrical	$55 < r < 116\,\text{cm}$
TID	3 disks	$58 < \|z\| < 124\,\text{cm}$
TEC	9 disks	$124 < \|z\| < 282\,\text{cm}$

transition region between barrel and endcap.

4.2 Electromagnetic Calorimeter

Outside the tracking system is the electromagnetic calorimeter (ECAL) [51]. The main role of the ECAL is to measure the energy of electromagnetic showers of electrons, photons, and neutral pions. It is a homogeneous and nearly hermetic calorimeter, made up of lead tungstate ($PbWO_4$) crystals, individually read by a photodetector. The ECAL consists of several subsystems: the ECAL barrel (EB), the ECAL endcap (EE) at the two sides, and a sampling ECAL preshower detector (ES) in front of EE with high granularity for discriminating neutral pions from prompt photons. The layout of the ECAL is shown in Figure 4.7.

Lead tungstate crystals have high density ($8.28\,\text{g/cm}^3$), short radiation length ($X_0 = 0.89\,\text{cm}$), small Moliére radius ($R_M = 2.2\,\text{cm}$), fast scintillation decay time and high radiation tolerance. These make them a suitable choice for a compact ECAL with fine granularity. Due to different magnetic fields and radiation levels, avalanche photodiodes (APD) and vacuum phototriodes (VPT) are chosen to be the photodetectors in the EB and EE, respectively.

The EB consists of a cylinder with an inner radius of 129 cm, a pseudorapidity

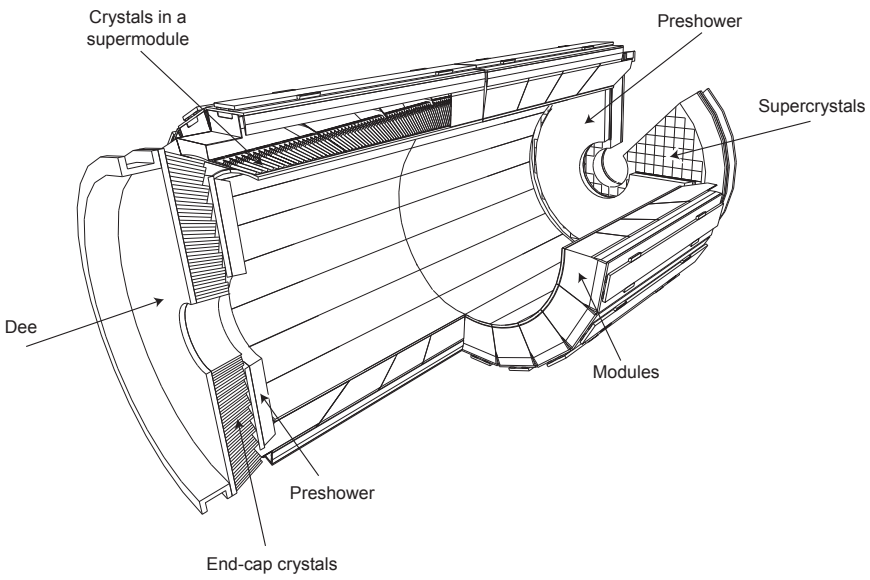


Figure 4.7: Layout of the CMS ECAL [52]. The ECAL barrel coverage is up to $|\eta| = 1.479$; the endcaps extend the coverage to $|\eta| = 3.0$; the preshower detector detector covers approximately $1.65 < |\eta| < 2.6$.

coverage up to $|\eta| = 1.479$, and a total number of 61200 $PbWO_4$ crystals. The crystals have a tapered shape: the cross section of the front face is about $22 \times 22\,\mathrm{mm}^2$, which translates into the $\Delta\eta \times \Delta\phi$ granularity of 0.0174×0.0174 ($1^\circ \times 1^\circ$); the cross section of the rear face is $26 \times 26\,\mathrm{mm}^2$; the total length is 230 mm, corresponding to 25.8 radiation lengths. The back of the crystals is attached to APDs for detecting the scintillation light from crystals.

The EE is located 3.14 m in the z direction from the nominal interaction point and covers the range of $1.479 < |\eta| < 3.0$. Each endcap contains 7324 $PbWO_4$ crystals. The crystals have a cross section of $28.6 \times 28.6\,\mathrm{mm}^2$ in the front and $30 \times 30\,\mathrm{mm}^2$ at the back, and a total length of 22 cm, corresponding to 24.7 radiation lengths. The VPTs are attached to the back of crystals for the photodetection. A summary of these information is provided in Table 4.2.

Table 4.2: Summary of ECAL parameters.

Parameter	Barrel	Endcap
Pseudorapidity coverage	$\|\eta\| < 1.479$	$1.479 < \|\eta\| < 3.0$
ECAL envelop: r_{inner}, r_{outer} [mm]	1238, 1750	316, 1711
ECAL envelop: z_{inner}, z_{outer} [mm]	0, $\pm$3045	$\pm$3170, $\pm$ 3900
Granularity: $\Delta\eta \times \Delta\phi$	0.0174×0.0174	0.02×0.02 to 0.05×0.05
Crystal dimension [mm^3]	21.8×21.8×230	28.6×28.6×220
Depth in X_0	25.8	24.7
No. of crystals	61200	14648

The relative energy resolutions in ECAL is parameterized with:

$$\frac{\sigma}{E} = \frac{S}{\sqrt{E}} \oplus \frac{N}{E} \oplus C\,, \tag{4.1}$$

where the three contributions are the stochastic, noise, and constant term. In 2006, the EB resolution was measured from an electron team beam by summing the energy in a grid of 3×3 crystals and found to be $S = 2.8\%$, $N = 12\%$, and $C = 0.3\%$ [53]. During the data-taking period, the transparency of the ECAL crystals decreases because of radiation damages, and then is partially recovered in the low-luminosity runs and technical stops. CMS uses a laser system continuously monitoring the crystal's light yield; the light response is corrected accordingly.

4.3 Hadron Calorimeter

Surrounding the ECAL is a hermetic brass/scintillator sampling hadronic calorimeter (HCAL) [54]. It is crucial for the measurement of neutral hadrons, since they do not create tracks in the tracker and leave only a small portion of their energy in the ECAL. With a hermetic structure, the HCAL plays an important role

for the reconstruction and measurement of jets and missing transverse momentum. The HCAL consists of four subsystems: the HCAL barrel (HB), the HCAL endcap (HE), the HCAL forward (HF) and the HCAL outer (HO). A layout of the HCAL with these different subsystems is shown in Figure 4.8.

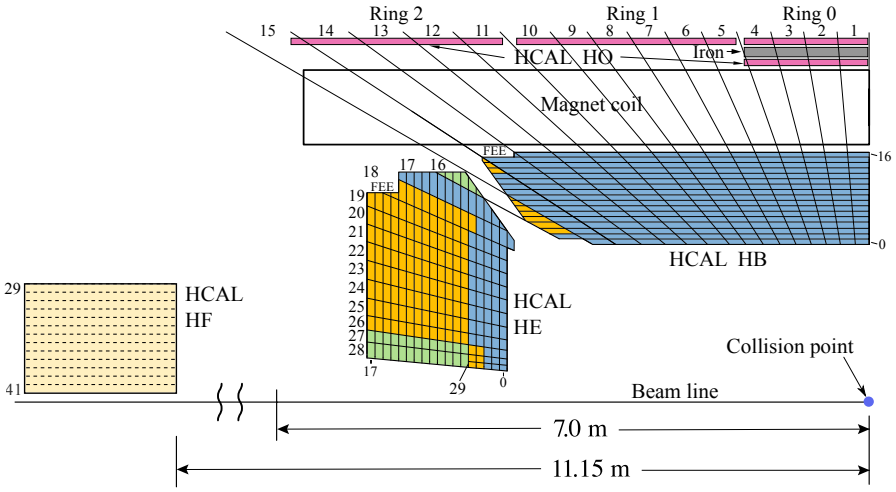


Figure 4.8: A schematic view of one quarter of the HCAL [55].

The HB is located between the EB (outer radius $r = 1.8\,\text{m}$) and the solenoid magnet (inner radius $r = 3.0\,\text{m}$), with the $|\eta|$ coverage up to 1.4. With the plane of $z = 0$ it is divided into two half barrels, each consisting of 18 identical wedges. Figure 4.9 shows the shape of an HB wedge as an example. Each wedge is segmented into 16 towers along η and 4 towers along ϕ, resulting in a tower size of 0.087×0.087 in $\Delta\eta \times \Delta\phi$ and a total number of 2304 towers in HB. There are 16 layers of absorber plates in each wedge, with 3.7 mm thick plastic scintillator tiles inserted in between (Layer 1 to Layer 15). The first plate is made of stainless steel with 40 mm thick, followed by eight 50.5 mm thick brass plates, six 56.5 mm thick brass plates, and a 75 mm thick steel plate for the last layer. The overall thickness of the HB absorber

ranges from about 6 nuclear interaction lengths at $\eta = 0$ to over 10 radiation lengths at $|\eta| = 1.3$. Before the first absorber layer and after the last absorber layer, there are also 9 mm thick scintillator layers installed (Layer 0 and Layer 16), with Layer 0 measuring the energy of hadronic showers in the dead material between the EB and the HB, and Layer 16 serving to correct for late developing showers leaking out of the HB. The light signals from plastic scintillators in different layers are collected by wavelength shifting (WLS) fibers and carried to hybrid photodiodes (HPDs) for signal digitizing and readout.

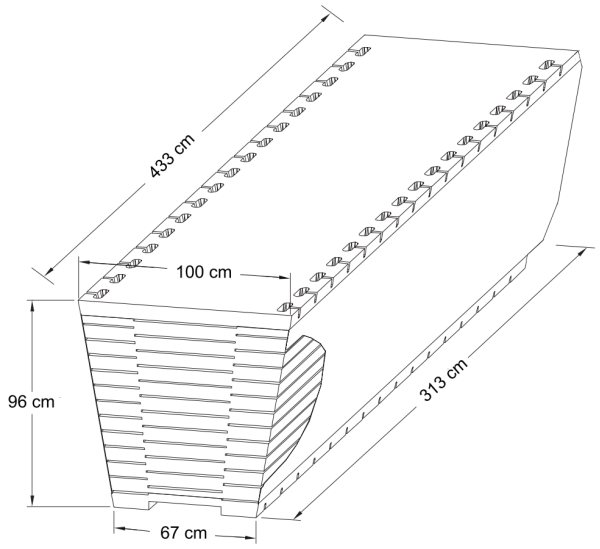


Figure 4.9: Isometric view of an HB wedge [56].

The thickness of HB is constrained by the outer radius of the EB and the inner radius of the solenoid magnet. Therefore, the HO detector is placed in the central region $|\eta| < 1.26$ outside the solenoid volume, to act as an extension of the HCAL system. The inclusion of the HO layers extends the total thickness of the calorimeter thickness to at least 11 nuclear interaction lengths.

The HE covers the range of $1.3 < |\eta| < 3.0$ and has a total number of 1368

towers. The granularity is the same as HB for $|\eta| < 1.6$, and becomes coarser at high $|\eta|$. The HE has a similar structure as the HB: a sampling detector with 78 mm thick brass absorber plates and 3.7 mm thick plastic scintillator tiles installed between the plates. It has an approximate thickness of 10 interaction lengths. In front of the first layer of absorber, 9 mm thick scintillators are installed for sampling from the dead material between the EE and the HE, similar to the HB case. Scintillators are connected with WLS fibers and HPDs are used for the readout. During the HCAL Phase-1 upgrade, all the HPDs in HB, HE and HO have been replaced with silicon photomultipliers (SiPMs) for higher radiation tolerance and better light yield resolution [57].

The HF is located outside the solenoid, at a distance of 11.2 m from the center interaction point. It covers the range of $2.85 < |\eta| < 5.2$ without any ECAL part in the front. Due to the much harsher radiation environment in the forward region, the HF design is different from the other HCAL subsytems. The HF consists of a steel absorber with a thickness of 165 cm (about 10 interaction lengths). Quartz fibers are embedded in the absorber as the active material and the produced Cherenkov light is read out by photomultiplier tubes (PMTs). The HF detector is divided into two longitudinal segments, with half of the fibers extending the full length of the absorber and the other half starting from a depth of 22 cm from the HF front face. This arrangement allows for the discrimination of electrons and photons from hadrons, making use of the fact that EM showers deposit most of their energy in the first 22 cm, while the hadronic showers deposit energy throughout the HF.

4.4 Muon System

The most outward part of the CMS detector is the muon system [58]. Unlike electrons, photons, or hadrons, muons are minimum ionizing particles and can fly through the detector mostly unaffected. The detection of a charged particle outside the magnetic solenoid is a clean and strong indicator that the particle is a muon. In CMS, the muon system is used to identify and trigger on muon objects, and also to improve the momentum resolutions of high p_{T} muons ($p_{\mathrm{T}} > 200\,\mathrm{GeV}$). Three types of gas ionization chambers were chosen to build the CMS muon system: drift tube chambers (DTs) in $|\eta| < 1.2$, cathode strip chambers (CSCs) in $0.9 < |\eta| < 2.4$, and resistive plate chambers (RPCs) in $|\eta| < 1.9$. All three subsystems are embedded in the steel flux-return yoke. A schematic view of the whole muon detector is shown in Figure 4.10.

In the barrel region, the neutron background is high, the magnetic filed is uniform, and therefore the DT chambers with long aluminum drift cells are used. There are 250 DT chambers in total, with each providing a spatial resolution of about $100\,\mu$m in r-ϕ. In the endcap region, because of the large neutron background and non-uniform magnetic field, CSC chambers are used. The spatial resolutions for CSC chambers are typically between 50-140 μm.

In order to complement the DTs and CSCs, RPCs are used in both the barrel and endcap regions up to $|\eta| = 1.9$. The RPCs have a coarse position resolution but an excellent intrinsic timing resolution of about 1.5 ns. They are mainly used to provide accurate measurements of the timing information of the muon hits with

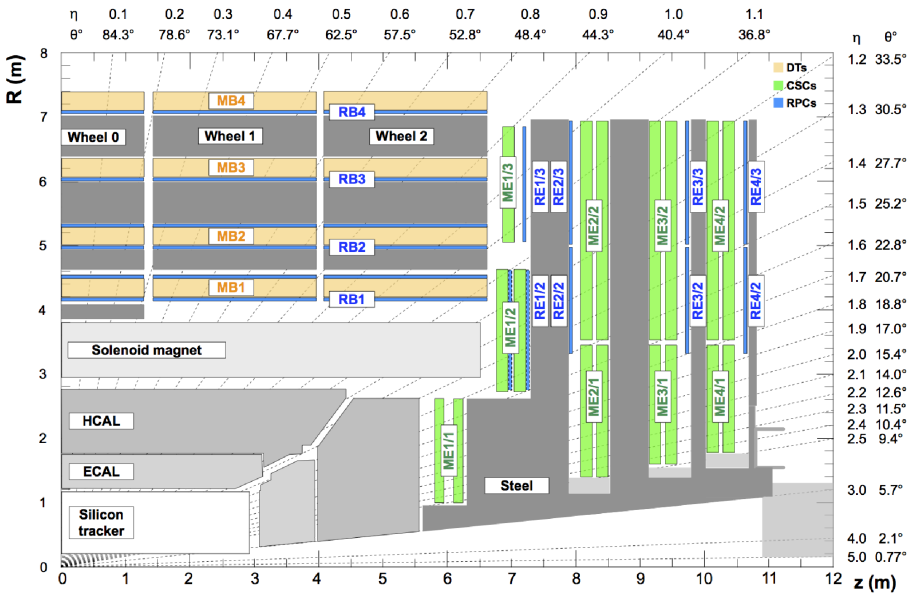


Figure 4.10: An r-z cross section of a quadrant of the CMS detector, including the muon subsystems: DTs, CSCs, and RPCs [59]. The DTs are labeled MB ("Muon Barrel") and the CSCs are labeled ME("Muon Endcap"). RPCs are mounted in both the barrel and the endcap, where they are labeled RB and RE respectively. The magnet yoke is represented by the dark gray areas.

acceptable position resolutions. The measurements are then combined with the precise position information from DTs and CSCs in order to trigger on a muon candidate, and assign it to the correct bunch crossing.

4.5 Trigger

As described in Chapter 3, the LHC operates with a bunch spacing of 25 ns, which translates into a 40 MHz bunch crossing frequency. (Due to the unfilled bunch gaps in between, the actual bunch crossing frequency is about $2556 \times 11245 = 28.7$ MHz.) Given there are multiple pp interactions per bunch crossing, the total pp inelastic collision rate is much higher, on the order of 1 billion per second. Among these only 1000 events per second can be saved due to the constraints on

the readout, storage, and computing. In CMS, a two-level trigger system [60–62] is utilized in order to select the events of potential physics interest and rejecting the non-interesting ones.

The first level (L1) of the CMS trigger [63] is implemented with custom hardware processors. It processes information from the calorimeters and the muon system of every single bunch crossing, and reduces the event rate to 100 kHz by making a decision on whether or not to save each event.

Figure 4.11 shows the diagram of the L1 trigger system during Run 2. The calorimeter trigger consists of two layers: Layer 1 receives, calibrates, and sorts the local energy deposits ("trigger primitives") from the ECAL and HCAL; Layer 2 uses these calibrated trigger primitives to reconstruct and calibrate physics objects, such as electrons, jets, and energy sums. The muon trigger include three muon track finders (MTF) that reconstruct muons in the barrel (BMTF, $|\eta| < 0.9$), overlap (OMTF, $0.9 < |\eta| < 1.2$) and endcap (EMTF, $1.2 < |\eta| < 2.4$) regions of the detector; the reconstructed muons are sent to the global muon trigger for final muon selection. Finally, the global trigger collects the muons and calorimeter objects and executes the algorithms with different selection criteria ("trigger menu") in parallel for the final trigger decision.

Th events passing the L1 trigger selections are then sent to the high-level trigger (HLT), which utilizes a farm of several thousand commercial processors (CPUs and GPUs, referred to as "Event Filter Farm") and further reduces the event rate from 100 kHz to 1 kHz. The HLT uses the full precision of data from the detector and runs the physics object reconstructions and selections following different HLT

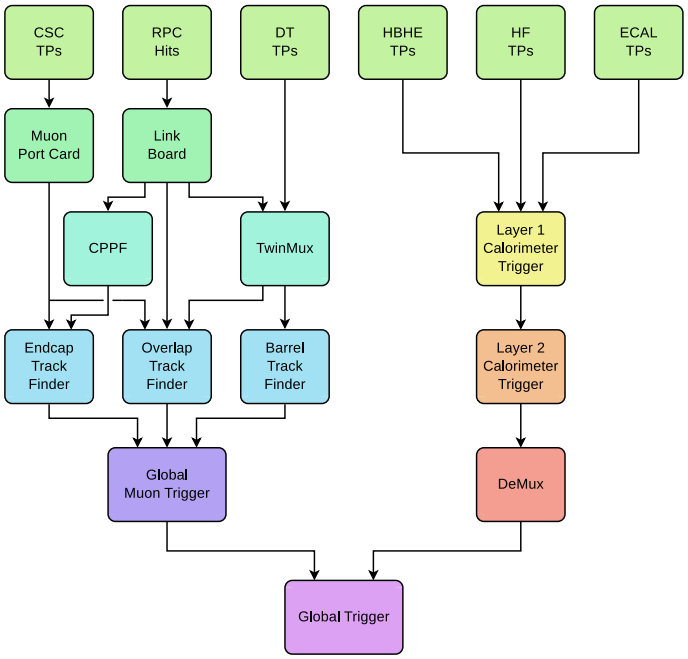


Figure 4.11: Diagram of the CMS Level-1 trigger system during Run 2 [63].

paths. Each HLT path is implemented as a sequence of steps of increasing complexity, reconstruction refinement, and physics sophistication. If one event fails the selection based on the calorimeter or the muon detector information, it will not be further processed such that the CPU-expensive tracking reconstruction will not be performed. The events passing the selections are then sorted into different datasets based on the HLT paths, and eventually transferred to the CMS Tier-0 computing center for offline processing and permanent storage.

In addition to the standard workflow described above, CMS has also developed other techniques to increase the event rate, such as data scouting and data parking [64]. The data scouting reduces the event size from the default of about 1 MB to around 1-10 kB, by running the event reconstruction during the online trigger processing and saving only the information of specific physics objects. The data parking sends the full raw events directly to tape without reconstruction. The

data are “parked” temporarily and reconstructed when the computing resources are available. These techniques allow for achieving an effective event rate of a few kHz.

Chapter 5: Object and Event Reconstruction

The physics object and event reconstruction needs to be carried out first in order to perform physics analyses. The reconstruction utilizes the different signatures that different types of particles create when they travel in the CMS detector. Figure 5.1 depicts these detection patterns. All charged particles, such as muons, electrons, and charged hadrons, create hits in the inner tracking system. Electrons and photons deposit most of their energy in the electromagnetic calorimeter. Charged and neutral hadrons deposit most of their energy in the hadronic calorimeter. Muons deposit little energy in the calorimeters, but create hits in the outer muon system. Neutrinos interact very weakly with the subdetectors, and their presence has to be inferred from the missing transverse momentum $p_{\mathrm{T}}^{\mathrm{miss}}$.

During the first step of the reconstruction process, digitized readouts from the subsystems of the CMS subdetectors are utilized to produce the reconstructed hits (*RecHits*), which contain information such as the position, energy deposition, and time of energy deposition. In the second step, the RecHits from different subsystems of one given subdetector are combined. Tracks and primary pp interaction vertices (*primary vertices*) are reconstructed from the tracker RecHits. Calorimeter towers (*CaloTowers*) are formed by summing over ECAL or HCAL RecHits from the same

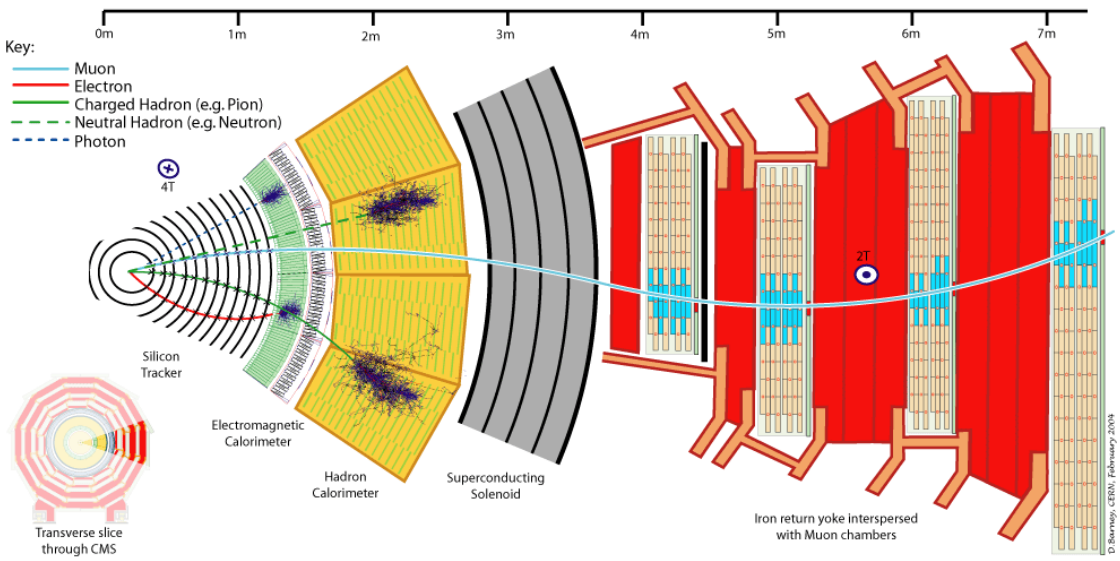


Figure 5.1: A sketch of how different particles interact with the CMS detector in the transverse slice [65].

tower. *Standalone* muons are created using the RecHits in the muon system. In the final step, these objects are correlated and the information from all subdetectors are combined, through the *particle-flow (PF) reconstruction* [66], to provide the optimal collection of physics objects with superior efficiencies and resolutions for physics analyses.

In this chapter, track and primary-vertex reconstruction is firstly introduced. Then a brief description of the PF reconstruction is presented, followed by a detailed description of muon, electron and photon reconstruction. Before the discussion of the jet and $p_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction, the pileup identification algorithm *pileup per particle identification* (PUPPI) [67] is discussed, which is widely used in the current CMS jet and $p_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction.

5.1 Tracks and Vertices

5.1.1 Track Reconstruction

A charged particle in a homogenous magnetic field follows a helical trajectory, which can be determined by 5 parameters. In the CMS coordinate system, they are $(d_0, z_0, \theta, \phi, q/p_{\mathrm{T}})$, defined at the point of closest approach of the track to the beam axis (*impact point*). d_0 and z_0 are the coordinates of the impact pact in the radial and z directions, respectively; θ is the polar angle; ϕ is the azimuthal angle; q/p_{T} is the ratio between the charge and the transverse momentum of the charged particle, also known as the signed reciprocal transverse momentum, from which the signed radius of the helix ρ can be determined with the known magnetic field B using $\rho = p_{\mathrm{T}}/qB$.

Track reconstruction [48] starts from the hits of charged particles in the pixel and strip detectors, and obtains their track parameters. In CMS the reconstruction is done with the Combinatorial Track Finder (CTF), which is an adaptation of the Kalman filter [68] that allows pattern recognition and tracking fitting to occur in the same framework.

The collection of reconstructed tracks is produced by iterating multiple times the CTF track reconstruction sequence (so-called "*iterative tracking*"). In the first step of each iteration, the track trajectories that are easiest to reconstruct (with large p_{T} and produced near the interaction point) are built by the track-finding module (*track finding*). Then the trajectories are fitted with the track-fitting module for

best estimations of the trajectory parameters (*track fitting*). Tracks that fail certain specified criteria, such as with large impact parameters, with many missing hits, or with large χ^2 from the track fitting, are discarded after the fitting (*track selection*). At the beginning of the next iteration, the hits associated with high quality tracks in previous iterations are removed, such that the combinatorial complexity is reduced and tracks that are more diffcult to reconstruct (with low p_{T}, or greatly displaced) can be built more efficiently.

Table 5.1 lists the requirements on the minimum p_{T} and the maximum transverse and longitudinal impact parameters relative to the center of the beam spot in different iterations. Figure 5.2 shows the track reconstruction efficiency as a function of the transverse distance from the beam axis to the production point of each particle. The efficiency is more than 90% for charged particles with $p_{\mathrm{T}} > 0.9\,\mathrm{GeV}$ produced within 2 cm from the beam axis. For tracks with 10-30 cm displacement in the transverse plane (outside the pixel tracker), the reconstruction efficiency is about 50-60%, and most of these tracks are reconstructed in the later iterations, as explained previously. The fake rate, defined as fraction of reconstructed tracks that are not associated with any simulated particle, is typically around a few percent.

The resolution of track p_{T}, transverse (d_0) and longitudinal (z_0) impact parameters depends on the track p_{T} and η, and other conditions, e.g., pileup. The p_{T} resolution is usually around a few percent; for tracks with p_{T} above 1 GeV, the d_0 resolution is usually around 20-30 μm, while for z_0 it is around 100 μm.

Table 5.1: List of requirements on the initial estimation of track trajectories for the track finding. d_0 and z_0 are the transverse and longitudinal impact parameters with respect to the beam spot. The σ in the table refers to the length of beam spot along the z direction, which is usually around 5 cm. The asterisk symbol indicates the z_0 in that case is calculated relative to a pixel vertex instead of to the center of the beam spot [48].

Iteration	$p_T >$ [GeV]	$d_0 <$ [cm]	$\lvert z_0 \rvert <$ [cm]
0	0.8	0.2	3σ
1	0.6	0.2	0.2^*
2	0.075	0.2	3.3σ
3	0.35	1.2	10
4	0.5	2.0	10
5	0.6	5.0	30

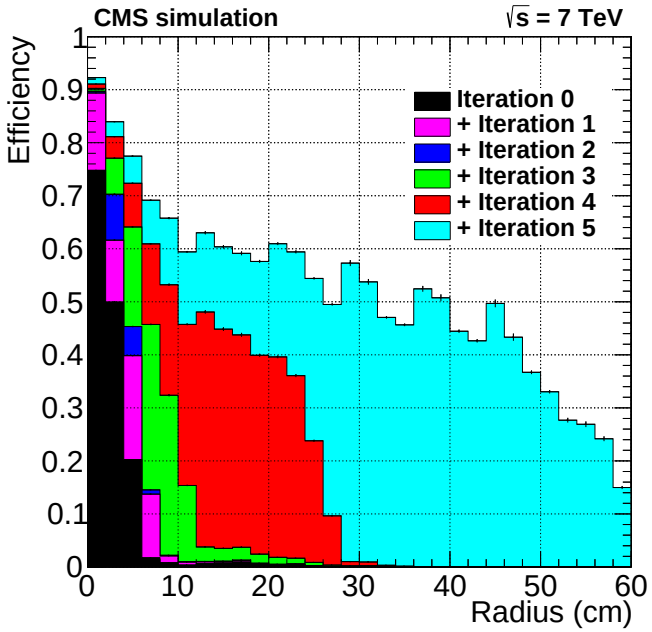


Figure 5.2: Track reconstruction efficiencies as a function of the transverse distance (r) from the beam axis to the production point of each particle, for tracks with $p_T > 0.9$ GeV and $|\eta| < 2.5$ [48].

5.1.2 Primary-vertex Reconstruction

The reconstructed tracks are used for the reconstruction of primary vertices, which include both the hard-scattering vertex and vertices from pileup collisions. Firstly, tracks are required to pass some quality cuts (e.g., number of hits in the pixel and strip tracker, and trajectory fit quality), and to be close to the interaction point (by cutting on the significance of the transverse impact parameter relative to the center of the beam spot). In the second step, selected tracks are clustered on the basis of their z coordinates at the point of closest approach to the center of the beam spot. The clustering is done using a *deterministic annealing* algorithm [69] to produce the candidate vertices. In the third step, the candidate vertices are fitted with an *adaptive vertex fitter* [70] for the best estimate of vertex parameters. Among all the reconstructed primary vertices, the one with the largest value of summed physics-object p_{T}^2 is selected as the hard scattering vertex (also called "leading vertex"). Other reconstructed primary vertices are referred to as pileup vertices.

Figure 5.3 shows the resolution of reconstructed primary vertices in the x and z directions. It is below 50 μm for vertices with more than 10 associated tracks.

5.2 Particle Flow

Particle Flow event reconstruction [66] is used to combine the measurements from all subdetectors for an optimal object event description. The reconstructed particles are different kinds of *PF candidates*: PF muons, PF electrons, PF photons,

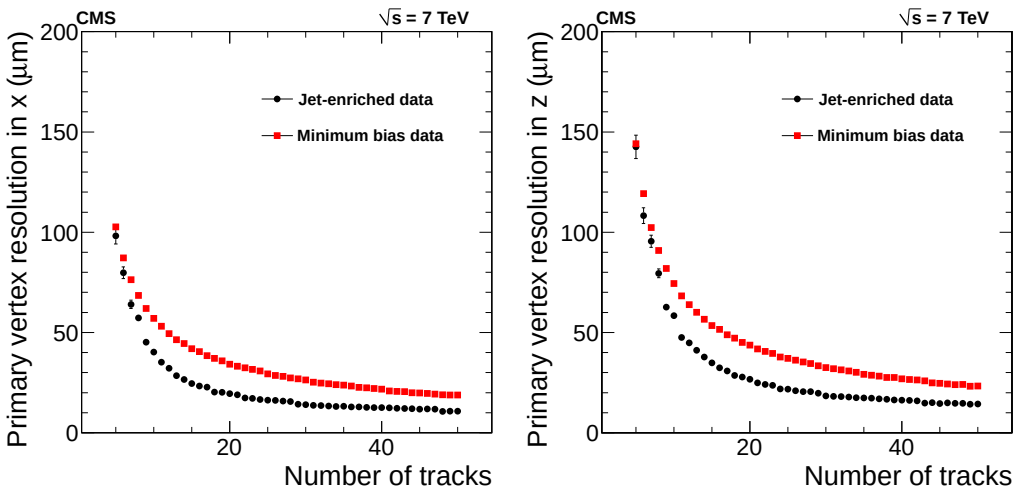


Figure 5.3: Resolution of reconstructed primary vertices in x (left) and z (right) directions, as a function of number of tracks associated with the vertex [48].

PF charged hadrons, PF neutral hadrons. In addition, the HF electromagnetic (EM) and hadronic (HAD) clusters are added to the particle list as HF photons and HF hadrons. No attempt is made to distinguish various species of neutral and charged hadrons in the PF reconstruction.

5.2.1 PF Elements and PF Blocks

The PF reconstruction relies on basic *PF elements*: reconstructed tracks from the inner tracker, calorimeter clusters in the ECAL and HCAL, and standalone muon tracks in the muon system. The calorimeter clustering is performed separately in EB, EE, HB, HE, and two PS layers. In the HF, no clustering is performed: the electromagnetic and hadronic components in each cell directly give rise to an HF EM and HAD cluster.

A *link algorithm* is applied to create links between PF elements from different subdetectors, such as a link between a track from inner tracker with a calorimeter

cluster, a link between HCAL clusters and ECAL clusters, a link between ECAL clusters and PS clusters, and a link between a track from inner tracker and a track from muon detector. The *PF blocks* are then produced, where one PF block represents one set of PF elements that are linked together, either by a direct link, or by an indirect link through common elements.

Figure 5.4 provides an example of the PF elements and PF blocks. There are five particles at the generator level: π^+, π^-, K^0_L, and two photons from one π^0 decay. Eight PF elements are reconstructed: the T_1 and T_2 tracks from the tracker, the $E_{1,2,3,4}$ clusters in the ECAL, and the $H_{1,2}$ clusters in the HCAL. Two links are produced: T_1 is linked with E_1 in the ECAL, and H_1 and H_2 in the HCAL; T_2 is linked with H_1 and H_2 in the HCAL; no additional link on other PF elements. Therefore, there are four PF blocks in total: T_1, T_2, E_1, H_1 and H_2 form a big PF block, since they are all linked together through the direct and indirect links; each of the $E_{2,3,4}$ clusters forms a unique PF block on its own, since it is not linked with other PF elements.

5.2.2 PF Candidates

PF candidates are identified and reconstructed from the PF elements in each PF block. The identification and reconstruction follows the order of PF muons, PF electrons, isolated PF photons, PF charged hadrons, and PF neutral hadrons and other PF photons. After the identification and reconstruction of each PF candidate, the corresponding PF elements will be removed from the PF block.

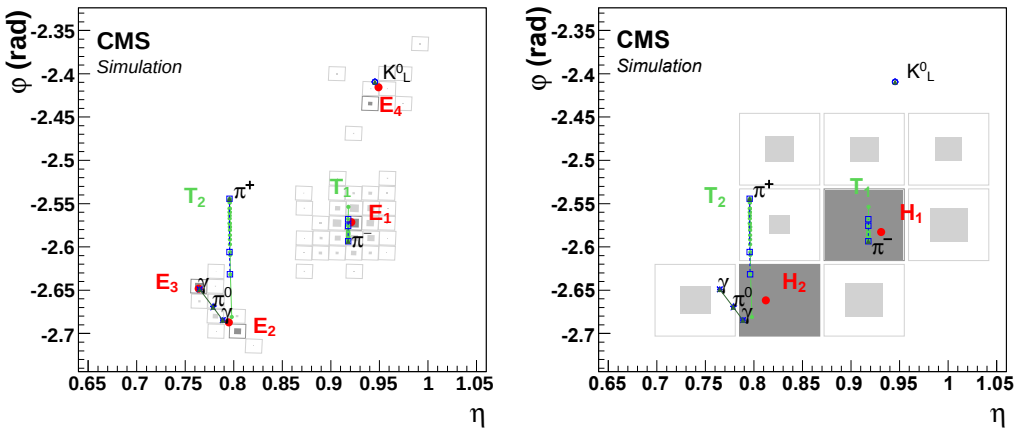


Figure 5.4: Event display of an illustrative jet made of five particles in the (η, ϕ) view on the ECAL (left) and HCAL (right) surface [66]. T_1 and T_2 are two charged tracks from π^- and π^+, respectively. The π^- also deposits its energy in the ECAL and HCAL, creating the E_1 cluster in the ECAL, and the H_1 and H_2 clusters in the HCAL. The π^+ creates no cluster in the ECAL, but two clusters H_1 and H_2 in the HCAL. The two photons from the π^0 decay create the E_2 and E_3 clusters in the ECAL. The K^0_L creates the E_4 cluster in the ECAL.

Muon candidates are identified and reconstructed in the first place. After that the corresponding PF elements (tracks and clusters) in the tracker, the calorimeters, and the muon system are removed from the PF block. Electrons and isolated photons are then identified and reconstructed, with corresponding tracks and clusters removed from the PF block afterwords as well.

The remaining tracks are then associated with charged hadrons. The track momenta are compared with the energy of their linked calorimeter clusters. If the calorimetric energy is significantly larger than the sum of the track momenta, the excess will be interpreted as PF photons and PF neutral hadrons.

The ECAL and HCAL clusters not linked to any track give rise to photons and neutral hadrons. Within the tracker acceptance ($|\eta| < 2.5$), all these ECAL clusters are associated with photons and all these HCAL clusters are associated

with neutral hadrons. Outside the tracker acceptance, ECAL clusters linked to a given HCAL cluster are assumed to arise from the same hadron shower, while ECAL clusters without such a link are associated with photons. These are based on the observation that, in hadronic jets, 25% of the jet energy is carried by photons, while neutral hadrons leave only 3% of the jet energy in the ECAL; while outside the tracker, charged and neutral hadrons are not distinguishable, and they leave 25% of the jet energy in the ECAL.

5.3 Muons

Muon tracks are reconstructed in the CMS detector in both the silicon tracker and the muon subdetectors (DT, CSC, and RPC), respectively resulting in *tracker tracks* and *standalone muons*. Subsequently, these tracks follow two reconstruction approaches [71].

The first approach is the *Global Muon reconstruction (outside-in)*. It starts from a standalone muon in the muon system and looks for a matched tracker track. If found out, a *global muon track* is fitted combining hits from the tracker track and standalone track.

The second approach is the *Tracker Muon reconstruction (inside-out)*. In this approach, all tracker tracks with transverse momentum $p_{\mathrm{T}} > 0.5\,\mathrm{GeV}$ and total momentum $p > 2.5\,\mathrm{GeV}$ are considered as possible muon candidates and are extrapolated to the muon system, taking into account the expected energy loss and the uncertainty due to multiple scattering. If at least one muon segment (i.e., a

short track stub made of DT or CSC hits) matches the extrapolated track, the corresponding tracker track qualifies as a *tracker muon track*.

At low momentum, the tracker muon reconstruction is more efficient than the global muon reconstruction, since it requires only a single muon segment in the muon system, but this increases the muon misidentification probability. Global muon reconstruction is designed to have high efficiency for muons penetrating through more than one muon station. By combining the information of the tracker and the muon system, the p_{T} resolution of global muons is also improved compared with the tracker muons, especially for muons with $p_{\mathrm{T}} > 200\,\mathrm{GeV}$, as shown in Figure 5.5.

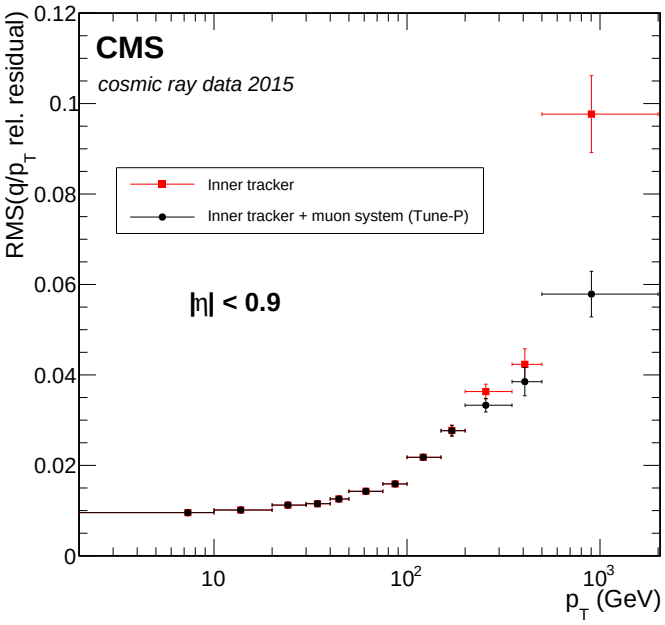


Figure 5.5: Resolutions of muon p_{T} as a function of cosmic muon p_{T}, using inner tracker fit only (red) and including the muon system (black) [71].

To distinguish prompt isolated muons with those from weak decays within jets, the muon isolation variable is defined. A cone of $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.4$ is built around the muon candidate, and the relative muon PF isolation is calculated

as:

$$I^{\mu}_{rel} = \frac{I_{ch,\mathrm{HS}} + max\left(I_{ph} + I_{nh} - 0.5 \cdot I_{ch,\mathrm{PU}}, 0\right)}{p^{\mu}_{\mathrm{T}}}, \tag{5.1}$$

where p^{μ}_{T} is the muon p_{T}; I_{ph}, I_{nh} are the p_{T} sum of all PF photons and PF neutral hadrons within the $\Delta R = 0.4$ cone of the muon candidate, respectively; $I_{ch,\mathrm{HS}}$ is the p_{T} sum of PF charged hadrons associated with the hard scattering vertex, and within the muon $\Delta R = 0.4$ cone; $I_{ch,\mathrm{PU}}$ is the p_{T} sum of PF charged hadrons associated with pileup vertices and within the $\Delta R = 0.4$ cone. The $0.5 \cdot I_{ch,\mathrm{PU}}$ term is the so-called "delta-beta" correction to correct for pileup contamination, and the factor 0.5 is estimated from simulations to be approximately the ratio of neutral to charged particles. The muon "tight" isolation requires $I^{\mu}_{rel} < 0.15$ [71]. Table 5.2 shows the full list of "tight" muon identification requirements.

Table 5.2: Full list of requirements for tight muon identifications.

Requirements	Tight
$p_{\mathrm{T}} >$ [GeV]	30
$\lvert\eta\rvert <$	2.4
Global	True
PF	True
$\chi^2/ndof <$	10
Valid muon hits $>$	0
Matched muon stations $>$	1
$d_{xy}(vtx) <$ [cm]	0.2
$d_z(vtx) <$ [cm]	0.5
pixel hits $>$	0
tracker layers with hits $>$	5
$I^{\mu}_{rel} <$	0.15
Tracker-based isolation	0.05

The reconstruction and tight identification efficiencies for muons with $p_{\mathrm{T}} > 20$ GeV are around 96% to 99%, with the misidentification rates of pions and kaons less than

0.5%. The efficiency for the tight PF isolation is more than 95% [71, 72].

5.4 Electrons

Electrons in the CMS detector are reconstructed via the association of a track from the silicon detector with a cluster of energy in the ECAL [73].

The electron energy is usually spread out over several crystals in the ECAL. These energy depositions undergo two steps of clustering. The first step starts from the "seed" crystals (the one contains the maximum energy deposit in a certain region) and groups clusters from continuous crystal arrays of 5×1 in $\eta \times \phi$ (the "hybrid" algorithm) for the ECAL barrel, or 5×5 crystals for the ECAL endcap (the "multi-5×5" algorithm). In the second step a supercluster (SC) comprising the energy of constituent clusters is formed, and the energy-weighted mean of the cluster positions is calculated as the SC position.

Electron tracks are formed from initial seeds likely to correspond to initial electron trajectories, which are then used to build tracks by collecting hits in the silicon tracker using the combinatorial Kalman filter procedure. Next, a track fitting procedure is undertaken using a Gaussian Sum Filter (GSF), in which the energy loss in each tracker layer is approximated by a mixture of Gaussian distributions. The supercluster position and energy, along with the GSF track, reconstruct the electron in the detector.

The electron momentum p_{comb} is estimated by combining the ECAL SC energy with the track momentum, through a regression technique used to define a weight

w such that $p_{comb} = wp + (1 - w)E_{SC}$. At low energies ($E < 15\,\text{GeV}$) or for electrons near gaps in the detector, the track momentum is expected to be more precise and contribute more to the combined momentum estimation. Figure 5.6 shows the effective resolution of electron momentum as a function of the generated energy for electrons in the barrel. The resolution is about 2% for barrel electrons after combining the ECAL and track information.

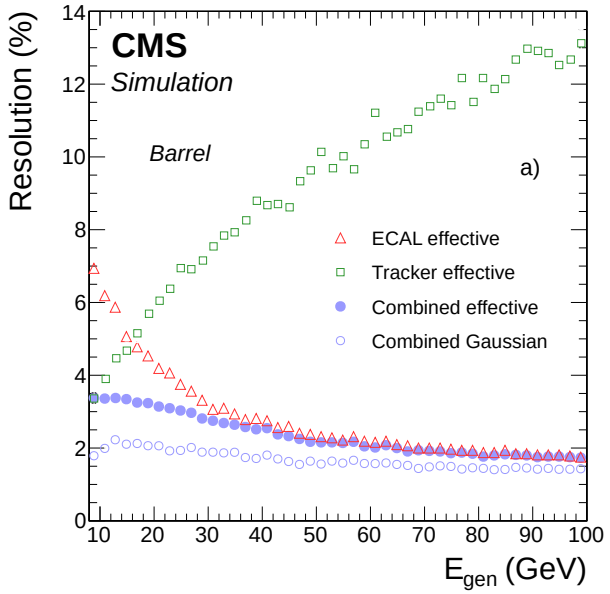


Figure 5.6: Effective resolutions of electron momentum as a function of the generated energy for electrons in the barrel [73].

The main backgrounds for the identification of prompt isolated electrons (signal) are photon conversions, jet misidentified as electrons, and non-isolated electrons from semileptonic decays inside heavy flavor jets. Quality cuts classified in several categories are utilized to distinguish signals from backgrounds.

1. Purely calorimetric observables. For example, the transverse shape variable of the electromagnetic showers $\sigma_{i\eta i\eta}$ is exploited given the fact that EM showers

are narrower than hadronic showers; the energy fractions deposited in the HCAL and ECAL are utilized because electrons deposit much larger fractions of energy in the ECAL than in the HCAL.

2. Purely tracking observables, for example, the missing inner hits in the tracker, the impact parameters in the transverse plane and in the longitudinal direction d_{xy} and d_z.

3. Observables checking the compatibility of the information from the ECAL and from the tracker, for example, the $\Delta\eta$ and $\Delta\phi$ between the SC and the track; the difference between the SC energy and track momentum $|1/E_{SC} - 1/p|$.

Similar to the muons, the isolation of electron candidates is computed from the flux of PF candidates found within a cone of $\Delta R = 0.4$ built around the electron direction. The flux of particles is computed independently for charged hadrons, neutral hadrons and photon candidates. When dealing with electron candidates, the neutral flux is corrected by using the average energy density due to pileup in the central region of the detector (ρ) and an effective area (A_{eff}) correction which normalizes this estimator in such a way that the isolation is independent of the number of pileup interactions. The electron isolation is therefore defined as:

$$I^e_{rel} = \frac{I_{ch} + max\left(I_{ph} + I_{nh} - A_{eff}\rho, 0\right)}{p^e_{\mathrm{T}}}\,. \tag{5.2}$$

5.5 Photons

Photons are reconstructed from superclusters in the ECAL in a similar manner as electrons [74].

As photons travel through the detector, they may convert into electron pairs. The R_9 variable is defined as the energy sum of the 3×3 crystals centered on the most energetic crystal in the SC divided by the energy of the SC. For photons converted before reaching the ECAL, the resulting showers have wider spread, and thus lower values of R_9 than those of unconverted photons. Figure 5.7 shows the photon energy resolutions in simulated $H \to \gamma\gamma$ events, as a function of $|\eta|$, for photons with $R_9 \geq 0.94$ and $R_9 < 0.94$, after the cluster energy corrections and calibrations. The resolution is about 1-2% in the barrel and 3% in the endcap.

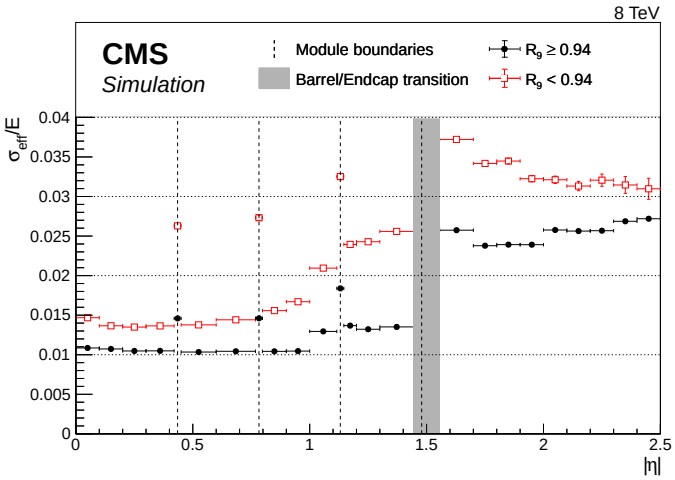


Figure 5.7: Photon relative energy resolutions as a function of $|\eta|$, in simulated $H \to \gamma\gamma$ events [74].

Similar to the electron identifications, the shower shape variable $\sigma_{i\eta i\eta}$, the energy deposited in the HCAL over the energy deposited in the ECAL H/E, and

the pileup corrected isolation variable are used for photon identifications. In order to reject electrons from photons, two more veto criteria are developed using the track information and are often applied to photon candidates in physics analyses. The first is called "conversion safe electron veto" (CSEV). The CSEV requires that no track should point to the cluster of the photon candidate in the ECAL, for the tracks that have at least one pixel hit and are not associated with a conversion vertex. The second is called "pixel seed veto" (PSV). The PSV is tighter than the CSEV and it requires that in the back propagation from the photon supercluster to the primary vertex, there should be no track seed identified in the pixel detector. This cut removes some converted photons, but also reduces the more electron background.

5.6 The PUPPI Algorithm

Pileup collisions can affect significantly the reconstruction of various physics objects and therefore pileup mitigation is needed for effective physics performance. In CMS, the widely used techniques include charged-hadron subtraction (CHS) [66], pileup jet identification [75], the "delta-beta" correction mentioned in Section 5.3, and pileup per particle identification (PUPPI) [67]. Among these, the PUPPI algorithm is found out to have excellent performance in various areas, such as the reconstruction of jets and missing transverse momentum, the lepton isolation, and the calculation of jet substructure observables for boosted object tagging [76].

The PUPPI algorithm exploits tracking information, local particle distribution, and event pileup properties in order to assign an individual weight for each

particle (in our case, to each PF candidate). The weights are in the range of 0 to 1 and represents the degree to which PF candidates are likely to be produced from the leading vertex (LV). When applying the PUPPI algorithm in object reconstructions, the momentum of each PF Candidate is rescaled with its weight accordingly.

The PUPPI algorithm starts with charged PF Candidates and assigns each of them a binary weight (either 1 or 0) utilizing the candidate's track-vertex association information. Charged PF candidates associated with the LV is assigned a weight of 1. If the charged PF candidate is not associated with any vertex, and its impact parameter in the longitudinal direction (d_z) is smaller than 0.3 cm, it get assigned a weight of 1 as well. In other scenarios, a weight of 0 is applied.

For neutral PF candidates, a local shape variable α is defined using the surrounding PF candidates. For a given candidate i, the α is defined as:

$$\alpha_i = \log \sum_{j \neq i, \Delta R_{ij} < 0.4} (\frac{p_{\mathrm{Tj}}}{\Delta R_{ij}})^2 \begin{cases} \text{for } |\eta_i| < 2.5,\ j \text{ are charged PF candidates from LV} \\ \text{for } |\eta_i| > 2.5,\ j \text{ are all kinds of reconstructed PF candidates} \end{cases} \tag{5.3}$$

where $\Delta R_{ij} = \sqrt{(\Delta\eta_{ij})^2 + (\Delta\phi_{ij})^2}$ is the distance between candidate i and candidate j in η-ϕ space, and the sum of j is over neighboring candidates within a cone of radius 0.4 in η-ϕ space. A large α_i value represents the PF candidates surrounding the candidate i are hard and central, implying the candidate i is likely to originate from the LV.

To convert the local shape variable α into the probability of originating from the LV, a few more variables are defined and described below. Firstly, the α values

for charged PF candidates from pileup (PU) are also calculated using the same definition, and the median and root-mean-square (RMS) of the α values, $\bar{\alpha}_{\mathrm{PU}}$ and $\alpha_{\mathrm{PU}}^{\mathrm{RMS}}$, are extracted for quantifying the pileup distribution in one event. Then, a signed χ^2 approximation is calculated for each neutral PF candidate,

$$\text{signed } \chi_i^2 = \frac{(\alpha_i - \bar{\alpha}_{\mathrm{PU}})|\alpha_i - \bar{\alpha}_{\mathrm{PU}}|}{(\alpha_{\mathrm{PU}}^{\mathrm{RMS}})^2}\,. \tag{5.4}$$

A large χ_i^2 represents a large α_i, and therefore the candidate i is likely to originate from the LV. In the last step, the signed χ^2 is transformed into a weight using:

$$w_i = F_{\chi^2,\mathrm{NDF}=1}(\chi_i^2)\,, \tag{5.5}$$

where $F_{\chi^2,\mathrm{NDF}=1}$ is the cumulative distribution function of a χ^2 with one degree of freedom.

Figure 5.8 shows the α, signed χ^2 and PUPPI weights distributions, where most charged particles associated with the pileup vertices are assigned small signed χ^2, as expected. Good data-MC agreement is achieved in the PUPPI weights calculation.

5.7 Jets

Jets [77] are collimated sprays of energetic hadrons arising from the fragmentation and hadronization of a quark or gluon. In CMS, jets are reconstructed from PF candidates using the anti-k_{T} clustering algorithm [78]. The algorithm is a specific

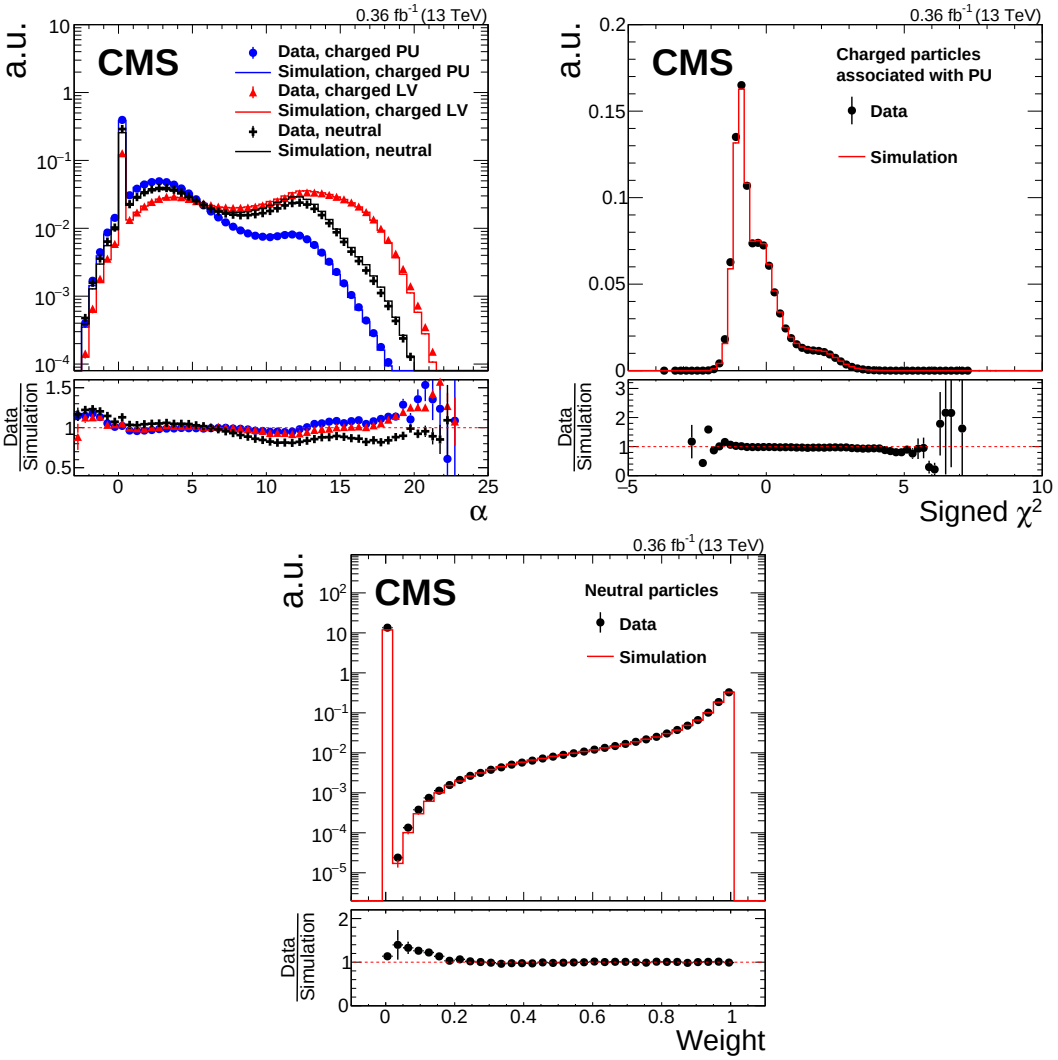


Figure 5.8: Data-MC Comparisons of three variables in PUPPI algorithm. The top left is the α distribution calculated using Equation 5.3. The top right is the distribution of the signed χ^2 from Equation 5.4. The bottom is the PUPPI weight distribution for neutral candidates. The markers show a subset of data taken in 2016, while the solid lines are QCD multijet simulations [76].

case of sequential clustering algorithms, described as

$$d_{ij} = \min(p_{\mathrm{T},i}^{2p}, p_{\mathrm{T},j}^{2p}) \frac{\Delta_{ij}^2}{R^2}\,, \tag{5.6}$$

$$d_{iB} = p_{T,i}^{2k}\,, \tag{5.7}$$

where $\Delta_{ij} = \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}$, and y_i and ϕ_i are the rapidity and azimuth angle of particle i, respectively; R is the radius parameter; d_{ij} defines the distance between particle i and particle j, while d_{iB} defines the distance between particle i and the beam.

In one iteration, both d_{ij} and d_{iB} are calculated for all particles. If the minimum distance is d_{iB}, then particle i is considered to be a fully-clustered jet and removed from the list of particles. Otherwise, particle i and particle j from the minimum d_{ij} are grouped together in the particle list, to be used in the next iteration. The iteration continues until the list is empty.

If $p = 1$, one recovers the inclusive k_{T} algorithm [79]. For $p = 0$, it becomes the Cambridge-Aachen (CA) algorithm [80, 81]. For $p = -1$, it is the anti-k_{T} algorithm. The so-called "AK4" jet reconstruction in CMS is implemented in FastJet version 3.0.1 [82], with the anti-k_{T} algorithm and the radius parameter $R = 0.4$.

Two approaches have been taken to mitigate the pileup effects on jets. The first is referred to as charged-hadron subtraction (CHS), as briefly mentioned in the previous section. In this approach, charged hadrons associated with vertices other than the LV are removed from the list of PF candidates for the jet reconstruction. Then jet momentum corrections are applied to remove the remaining energy from

neutral and charged particles from pileup. The jets reconstructed with this approach are so-called "CHS" jets. In the second approach, PUPPI weights are used to rescale the PF candidates p_{T}; the rescaled transverse momenta are used in the clustering. The jets reconstructed using PUPPI weights are so-called "PUPPI" jets; no extra event-by-event pileup corrections are needed for these jets.

The jet energy is calibrated sequentially after the reconstruction. The calibration follows the sequence of pileup offset subtraction, detector response correction from simulation, and residual corrections for the differences between data and simulation [83].

A set of variables are applied to identify jets from various backgrounds. These variables include the charged EM and hadron energy fractions, the neutral EM and hadron energy fractions, number of associated PF candidates, number of associated charged and neutral candidates, and muon energy fractions. The cuts applied on these variables ensure the least contamination from pileup and leptons, while maintaining a high identification efficiency [75].

5.8 Missing Transverse Momentum

According to momentum conservation, the net transverse momentum of all final-state particles from pp collisions should balance to zero. The missing transverse momentum (MET) $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$, which is the momentum imbalance in the transverse plane, and its magnitude $p_{\mathrm{T}}^{\mathrm{miss}}$, can be used to infer the undetected weakly interacting neutral particles, such as neutrinos from SM W and Z boson decays and

hypothetical dark matter candidates from physics beyond the SM. Reconstruction and performance of $p_{\mathrm{T}}^{\mathrm{miss}}$ are sensitive to experimental resolutions, mismeasurements of reconstructed particles, detector artifacts and pileup interactions.

Currently there are two types of $p_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction algorithms widely employed in the CMS event reconstructions. The first one is PF $p_{\mathrm{T}}^{\mathrm{miss}}$, defined as the negative vector p_{T} sum of all PF candidates in the event [84–86]:

$$\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}} = -\sum_{i} \vec{p}_{\mathrm{T},i} \tag{5.8}$$

The second one, referred to as PUPPI $p_{\mathrm{T}}^{\mathrm{miss}}$, utilizes the PUPPI weights (w_i) information and calculates the $p_{\mathrm{T}}^{\mathrm{miss}}$ with the negative vector p_{T} sum of all PF candidates, weighted by their w_i:

$$\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}} = -\sum_{i} w_i \vec{p}_{\mathrm{T},i} \tag{5.9}$$

Compared with PF $p_{\mathrm{T}}^{\mathrm{miss}}$, PUPPI $p_{\mathrm{T}}^{\mathrm{miss}}$ is less pileup dependent and has smaller resolutions.

Detector effects can lead to an inaccurate estimation of $p_{\mathrm{T}}^{\mathrm{miss}}$, for example, the nonlinearity of the calorimeter response to hadrons, and the minimum energy (p_{T}) thresholds for neutral (charged) particle reconstructions. The so-called "Type-I" corrections are applied to improve the $p_{\mathrm{T}}^{\mathrm{miss}}$ estimation, where the jet p_{T} corrections ($\vec{p}_{\mathrm{T,jet}}^{\,\mathrm{corr}} - \vec{p}_{\mathrm{T,jet}}$) are propagated into the $p_{\mathrm{T}}^{\mathrm{miss}}$ calculation:

$$\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}} = \vec{p}_{\mathrm{T}}^{\,\mathrm{miss,raw}} - \sum_{\mathrm{jets}} (\vec{p}_{\mathrm{T,jet}}^{\,\mathrm{corr}} - \vec{p}_{\mathrm{T,jet}}) \tag{5.10}$$

Performances of PF and PUPPI $p_{\mathrm{T}}^{\mathrm{miss}}$ will be discussed in detail in the next chapter. In this dissertation they are after the Type-I corrections by default.

Chapter 6: Introduction to Deep Neural Networks

Machine-learning techniques, especially modern deep-learning techniques, have been quickly adopted in various subfields in particle physics and have shown significant improvements compared with classical methodologies, as stated in Chapter 1. A basic introduction to deep neural networks and the relevant techniques adopted in the DeepMET training are briefly reviewed in this chapter. A more comprehensive review of deep-learning techniques can be found, for example, in [87–89].

6.1 Artificial Neuron and Neural Network

An artificial neural network is based on a collection of connected units called artificial neurons, which are analogous to biological neurons in biological brains. Figure 6.1 shows the basic structure of a single artificial neuron.

The pieces of information $\{x_1, ..., x_m\}$ from multiple sources, either directly from the input data, or from the outputs of previous neurons, are aggregated in the artificial neuron as a weighted sum z:

$$z = \sum_{i=1}^{m} w_i x_i + b\,, \tag{6.1}$$

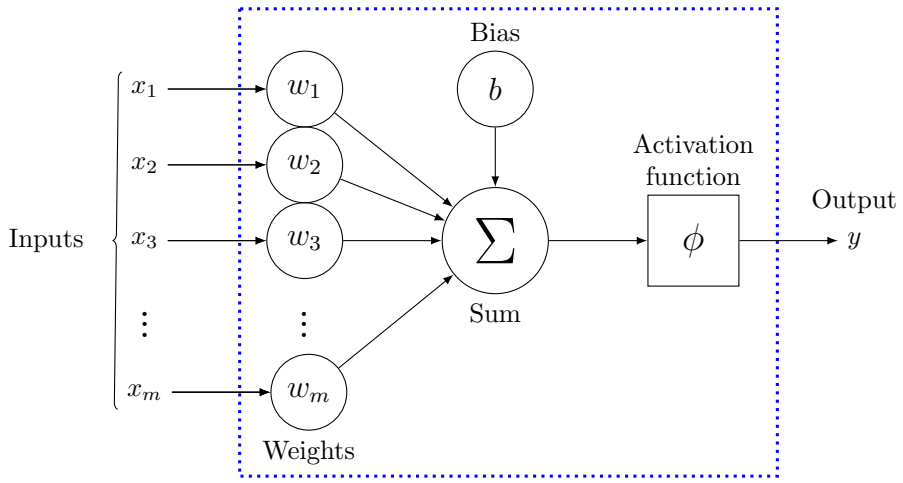


Figure 6.1: Illustration of the structure of a single artificial neuron.

where $\{w_0, w_1, ..., w_m\}$ and b are the weights and the bias stored in the neuron, respectively. The aggregated information z is then sent to a nonlinear function $\phi(z)$, often referred to as the "activation function", for further processing. The output of the neuron y is:

$$y = \phi(z) = \phi(\sum_{i=1}^{m} w_i x_i + b)\,. \tag{6.2}$$

In practice, the most commonly used activation functions include the Rectified Linear Unit (ReLU) [90]:

$$\mathrm{ReLU}(x) = \begin{cases} x, & x \geq 0 \\ \\ 0, & x < 0 \end{cases}, \tag{6.3}$$

the output of which is in the range of $[0, +\infty)$, and the sigmoid function:

$$S(x) = \frac{1}{1 + e^{-x}}\,, \tag{6.4}$$

the output of which is in the range of $[0, 1]$.

An n-dimensional dense layer is a layer of n such artificial neurons, which takes the m-dimensional input data $\boldsymbol{x}$, processes the information, and sends out an n-dimensional output $\boldsymbol{y}$:

$$\boldsymbol{y} = \phi(\boldsymbol{W}\boldsymbol{x} + \boldsymbol{b}), \tag{6.5}$$

where $\boldsymbol{W}$ is an $n \times m$ weight matrix, and $\boldsymbol{b}$ is an n-dimensional bias vector. A deep neural network is a network with multiple layers between the input and output layers ("hidden layers"). Figure 6.2 shows an illustration of a deep neural network. Deep neural networks can represent, according to the universal approximation theorem [91], a wide variety of interesting functions, by making use of an appropriate number of neurons and layers.

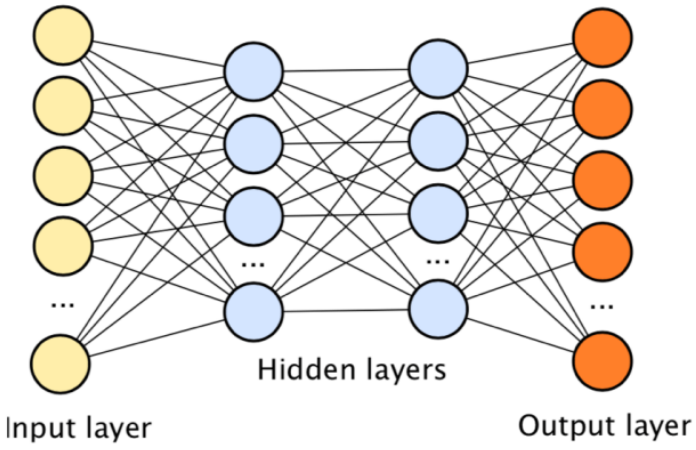


Figure 6.2: Illustration of the structure of a deep neural network.

The architecture of a deep neural network, such as the number of neurons and the activation function of each layer, is predefined (the "prior model") and can not be updated during the training. Instead, the elements in the weight matrix $\boldsymbol{W}$ and

the bias vector $\boldsymbol{b}$ are free parameters ("trainable parameters"). They are randomly initialized, and their values are tuned during the training phase in order to minimize the chosen loss function.

6.2 DNN Training and Loss Function

The DNN training is an optimization process analogous to a "fit" frequently performed in experimental particle physics. In a fit, a prior model is chosen with some free parameters $\boldsymbol{\theta}$. The goal of the fit is to optimize the parameters $\boldsymbol{\theta}$ such that the similarity between the predictions from the model $\boldsymbol{y}$ and the observations from the data $\hat{\boldsymbol{y}}$ is the largest. The χ^2 or the likelihood function is often chosen as the metric to quantify such similarity.

In machine learning, the prior model is replaced by a certain ML architecture with some trainable parameters $\boldsymbol{\theta}$, and the metric function chosen to be minimized is called the "loss function". (In the opposite case, the function chosen to be maximized is called the "reward function".) The most commonly used loss functions include the mean absolute error (MAE) and the mean square error (MSE) for continuous output variables:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} \left| \boldsymbol{y}_i - \hat{\boldsymbol{y}}_i \right| , \tag{6.6}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n} \left(\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i \right)^2 , \tag{6.7}$$

and the cross entropy loss function for binary classification output variables:

$$CE = -\sum_{i=1}^{n} [\boldsymbol{y}_i \log \hat{\boldsymbol{y}}_i + (1 - \boldsymbol{y}_i) \log(1 - \hat{\boldsymbol{y}}_i)] \,, \tag{6.8}$$

where $i = 1, 2, ..., n$ refers to i-th event.

The loss function is minimized during the training usually via the so-called "minibatch (stochastic) gradient descent" method. The whole training dataset is divided into a certain number of batches with a predefined batch size, which is typically between $\mathcal{O}(10)$ and $\mathcal{O}(1000)$. The total loss function of one batch, $\boldsymbol{L}(\boldsymbol{\theta})$, is calculated as the loss sum of all the events in that batch. The gradient of the total loss with respect to the trainable parameters $\boldsymbol{\theta}$ is calculated and the trainable parameters can be updated via:

$$\boldsymbol{\theta}_i' = \boldsymbol{\theta}_i - \alpha \frac{\partial}{\partial \boldsymbol{\theta}_i} \boldsymbol{L}(\boldsymbol{\theta}) \,, \tag{6.9}$$

where α is a pre-defined parameter (hyperparameter) and is referred to as the "learning rate". It controls how quickly the model is adapted in the training.

After all the batches are passed to the neural network once (one "epoch"), the loss function is evaluated on an independent dataset ("validation dataset") to validate the neural network performance with the updated trainable parameters. The ratio between the number of the training events and validation events is another hyperparameter, usually set to be in the range of $1:1$ and $10:1$.

In summary, these two section briefly describe an artificial neuron and a (deep)

neural network, with many hyperparameters introduced, such as the number of neurons, number of layers, the activation function, the batch size, the learning rate, and the training-validation split. These hyperparameters can have, in some cases, a significant impact on the training. The tuning of these parameters is mostly based on experience and experimental results, with some guidelines reported in [92].

6.3 Techniques to Improve the DNN Performance

Some techniques, frequently applied to the training of DNNs, have shown great improvements on the performance of DNNs. This section briefly describes two of them that are applied in the DeepMET architecture.

The first technique is Batch Normalization [93]. It consists in the normalization of each input using its mean value $\mu_{\mathcal{B}}$ and its standard deviation $\sigma_{\mathcal{B}}$ to reduce the covariate shifts and make the training converge faster. In a batch with size n, the mean value $\mu_{\mathcal{B}}$ and the standard deviation $\sigma_{\mathcal{B}}$ of the input x are calculated first:

$$\mu_{\mathcal{B}} = \frac{1}{n}\sum_{i=1}^{n} x_i \,, \tag{6.10}$$

$$\sigma_{\mathcal{B}}^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \mu_{\mathcal{B}})^2 \,. \tag{6.11}$$

The output of the Batch Normalization y is defined as:

$$y_i = \gamma \frac{1}{\sigma_{\mathcal{B}}}(x_i - \mu_{\mathcal{B}}) + \beta \,, \tag{6.12}$$

where γ and β are trainable scale and shift parameters.

The second technique consists in the definition of an Embedding layer [94]. It applies to categorical features, and transforms the input discrete numbers into vectors in a continuous latent space. For a categorical feature x_i with dimension m ($x_i = 1, 2, 3...m$), the embedding layer with n-dimensional output (n is a hyper-parameter) builds a "look-up table" that maps each x_i value to an n-dimensional embedding vector. The components of these vectors are trainable, and are optimized during the training process.

These two techniques are implemented in DeepMET, with the first making the training converge fast and the second improving the DeepMET performance. Their implementation and impact on DeepMET are described in detail in the following chapter.

Chapter 7: DeepMET

A new Deep-Neural-Network–based (DNN-based) $p_{\mathrm{T}}^{\mathrm{miss}}$ estimator, DeepMET, is presented in this chapter. Its performance has been observed to be consistently better than the PF and PUPPI $p_{\mathrm{T}}^{\mathrm{miss}}$ estimators commonly used in the CMS Collaboration.

The inputs to the DNN are the PF candidates reconstructed in each event. The DNN operates on each PF candidate individually, utilizes input features of each PF candidate, and assigns to each of them an individual weight (w_i) and two biases ($b_{i,x}, b_{i,y}$). The output of the estimator is defined as the negative of weighted p_{T} sum of all PF candidates together with their bias contributions:

$$p_x^{\mathrm{miss}} = -\sum_i (w_i p_{i,x} + b_{i,x}) \,, \tag{7.1}$$

$$p_y^{\mathrm{miss}} = -\sum_i (w_i p_{i,y} + b_{i,y}) \,, \tag{7.2}$$

similar to the PUPPI $p_{\mathrm{T}}^{\mathrm{miss}}$ calculation.

More details on the DeepMET estimator and its performance are presented in the following sections.

7.1 Parametrization of $p_{\mathrm{T}}^{\mathrm{miss}}$

The strategies and metrics that are typically used to evaluate the $p_{\mathrm{T}}^{\mathrm{miss}}$ performance are quickly reviewed in this section, before DeepMET is introduced.

The performance of $p_{\mathrm{T}}^{\mathrm{miss}}$ is usually studied in samples with a well-identified Z boson decaying into a pair of electrons or muons, or with a well-identified high-p_{T} isolated photon [86]. According to the momentum conservation in the transverse plane, the p_{T} of the vector boson, the $p_{\mathrm{T}}^{\mathrm{miss}}$, and the hadronic recoil in one event should follow:

$$\vec{q}_{\mathrm{T}} + \vec{u}_{\mathrm{T}} + \vec{p}_{\mathrm{T}}^{\mathrm{miss}} = 0\,, \tag{7.3}$$

where $\vec{q}_{\mathrm{T}}$ is the $\vec{p}_{\mathrm{T}}$ of the Z boson or the photon; $\vec{u}_{\mathrm{T}}$ is the hadronic recoil of the vector boson in the transverse plane, calculated as the vector sum of all reconstructed PF candidates except for the vector boson.

The performance of $p_{\mathrm{T}}^{\mathrm{miss}}$ can thus be measured from u_{T} and q_{T} based on Equation 7.3. The boson p_{T}, q_{T}, is precisely measured either directly from a high-p_{T} photon, or two high-p_{T} leptons from the Z decay. The hadronic recoil u_{T} instead suffers from uncertainties on the soft activities such as pileup, and calorimeter noises and resolutions. The measurement of q_{T} is therefore significantly better than u_{T}, resulting in the $p_{\mathrm{T}}^{\mathrm{miss}}$ performance to be largely dominated by the u_{T} contributions. The response and resolution of $p_{\mathrm{T}}^{\mathrm{miss}}$ can be studied from the response and resolution of the hadronic recoil u_{T}.

Little or no genuine $p_{\mathrm{T}}^{\mathrm{miss}}$ exists in Z/γ + jets events, leading to the approxi-

mation:

$$\vec{q}_{\mathrm{T}} + \vec{u}_{\mathrm{T}} = 0\,, \tag{7.4}$$

in the transverse plane. As illustrated in Figure 7.1, $\vec{u}_{\mathrm{T}}$ can be decomposed into two components: $u_{\parallel}$ and $u_{\perp}$, parallel and perpendicular to the boson axis, respectively. A good $p_{\mathrm{T}}^{\mathrm{miss}}$ (u_{T}) estimator should have:

- the response close to unity, $-\langle u_{\parallel}\rangle/\langle q_{\mathrm{T}}\rangle = 1$;
- the resolution of $u_{\parallel}$ and $u_{\perp}$, denoted by $\sigma(u_{\parallel})$ and $\sigma(u_{\perp})$, as small as possible.

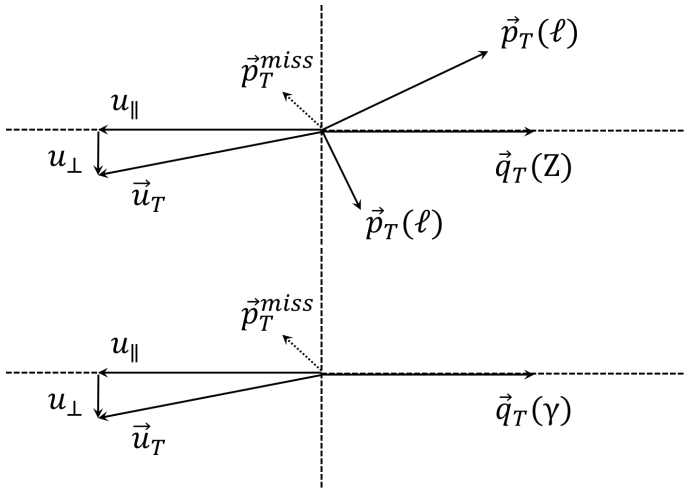


Figure 7.1: Illustration of parametrization of the Z boson (top) and photon (bottom) event kinematics in the transverse plane. Such events have little or no genuine $p_{\mathrm{T}}^{\mathrm{miss}}$ [86].

In the following sections, unless otherwise stated, the resolution of $u_{\parallel}$ ($u_{\perp}$) is defined as:

$$\sigma = \frac{1}{2}(q_{84} - q_{16})\,, \tag{7.5}$$

where q_{84} and q_{16} are the 84% and 16% quantiles of the $u_{\parallel}$ ($u_{\perp}$), respectively. This metric is more robust to outliers than the RMS.

To summarize this section, the performance of $p_{\mathrm{T}}^{\mathrm{miss}}$ is studied in $\mathrm{Z}(e^+e^-/\mu^+\mu^-)$ + jets or γ + jets events, where there is little or no genuine $p_{\mathrm{T}}^{\mathrm{miss}}$. The response and resolution of $p_{\mathrm{T}}^{\mathrm{miss}}$ are measured from the response and resolution of the hadronic recoil u_{T}. The response, which should be close to unity, is measured from $-\langle u_{\parallel}\rangle/\langle q_{\mathrm{T}}\rangle$, while the resolution, expected to be as small as possible, is measured from $\sigma(u_{\parallel})$ and $\sigma(u_{\perp})$.

7.2 Datasets and Event Selections

7.2.1 Data and MC simulations

The data analyzed in this dissertation consist of $35.9\,\mathrm{fb}^{-1}$ of integrated luminosity, collected by the CMS detector during the 2016 run of the LHC, with all CMS subdetectors operating within their nominal conditions. The events are selected with single-muon triggers, with the online muon p_{T} threshold of 24 GeV at the HLT. The average pileup in the data sample is 23.

Simulated Monte Carlo (MC) events are used to model the $\mathrm{W}(\ell\nu)$+jets, $\mathrm{Z}(\ell\ell)$+jets, and other SM processes. The $\mathrm{W}(\ell\nu)$+jets and $\mathrm{Z}(\ell\ell)$+jets samples are simulated with up to 2 jets at next-to-leading order (NLO) in perturbative QCD with the MadGraph5_aMC@NLO version 2.2.2 event generator [95]. The simulation of these processes with the same generator and up to 4 jets at leading order (LO) is also used for cross validations. Other relevant processes are simulated with MadGraph5_aMC@NLO ($\mathrm{t\bar{t}}$ with up to 3 jets at LO, $\mathrm{WZ}\to 3\ell\nu$ with up to 1 jet at NLO), as well as with POWHEG 2.0 [96–98] at NLO ($\mathrm{WW}\to 2\ell 2\nu$, $\mathrm{ZZ}\to 2\ell 2\nu$,

ZZ$\rightarrow 4\ell$).

All simulated samples are interfaced with the PYTHIA 8.226 [99] package and the CUETP8M1 [100] tune for parton showering, hadronization, and underlying event simulation. For the MADGRAPH5_aMC@NLO samples, jets from the matrix element calculations are matched to the parton shower following the MLM [101] (FxFx [102]) prescription for LO (NLO) samples. The NNPDF3.0 [103] set of parton distribution functions (PDFs) are used for all samples, with the order matching the order of the matrix element calculations. The simulation of the interactions of all final-state particles with the CMS detector is done with GEANT4 [104]. Pileup effects are also added to each simulated event sample.

The full list of used MC samples and their cross sections are reported in Table 7.1.

Table 7.1: List of used MC simulation samples and their cross sections. Each dataset name is followed by */RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v*/MINIAODSIM*.

Dataset Name	Cross Section [pb]
/WJetsToLNu_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	61526.7
/WJetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	61526.7
/DYJetsToLL_M-50_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	6077.2
/DYJetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	6077.2
/TTJets_DiLept_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	87.3
/TTJets_SingleLeptFromT_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	182.2
/TTJets_SingleLeptFromTbar_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	182.2
/WZTo3LNu_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	5.26
/WWTo2L2Nu_13TeV-powheg	12.2
/ZZTo2L2Nu_13TeV_powheg_pythia8	0.56
/ZZTo4L_13TeV_powheg_pythia8	1.21

7.2.2 Event Selections

In this dissertation most of the studies are done in the μ channel: the $Z \to \mu^+\mu^-$ and $W \to \mu\nu$ decays. The events from the electron channel have similar properties, and similar performances have been confirmed in other studies. The total number of events in these channels are large: $\mathcal{O}(10^9)$ $W(\mu\nu)$ events and $\mathcal{O}(10^8)$ $Z(\mu\mu)$ events. The selection criteria are well established, and thus limited effort has been put in the optimizations of the kinematic requirements, which are based on the SingleMuon trigger. Detailed selection criteria for different targeted processes are provided in the following contents.

To pass the single-muon selections, the event is required to satisfy the following criteria:

- the event should pass the single muon trigger `HLT_IsoMu24` OR `HLT_IsoTkMu24`;
- the leading muon should have $p_T > 30\,\text{GeV}$ and $|\eta| < 2.4$, and pass the tight muon identification requirements, with the relative muon PF isolation from Equation 5.1 less than 0.15;
- no additional muon with $p_T > 10\,\text{GeV}$ and passing the tight muon identification requirements should be present; no electron with $p_T > 10\,\text{GeV}$ and passing the medium electron identification requirements should be present.

To pass the dimuon selections, the event is required to satisfy the following criteria:

- the event should pass the muon trigger `HLT_IsoMu24` OR `HLT_IsoTkMu24`;

- the leading muon should have $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 2.4$, and pass the tight muon identification requirements, with the relative muon PF isolation less than 0.15;

- the subleading muon should have $p_{\mathrm{T}} > 25$ GeV and $|\eta| < 2.4$, and pass tight muon identification requirements, with the relative muon PF isolation less than 0.15.

Figure 7.2 shows the distribution of the invariant mass of the two muons after the dimuon selections, which demonstrates a relatively good data-MC agreement.

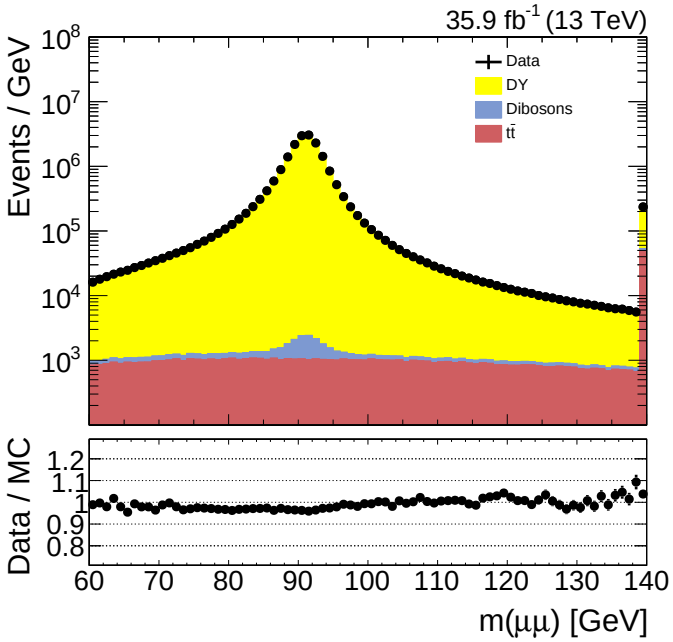


Figure 7.2: Distribution of the invariant mass of the two muons after the dimuon selections. The overflow contents are included in the last bin.

Events with genuine $p_{\mathrm{T}}^{\mathrm{miss}}$ need to be vetoed in order to validate the $p_{\mathrm{T}}^{\mathrm{miss}}$ performance in data. Such events are dominated by the $\mathrm{t\bar{t}}$ process in the dimuon final states. Therefore the $\mathrm{Z} \to \mu\mu$ selections includes the following requirements:

- the event should pass the dimuon selections listed above;
- the invariant mass of the two muons $m(\mu\mu)$ in the event should be in the range: $|m(\mu\mu) - 91.0| < 10.0\,\text{GeV}$;
- b jet veto: for all the jets passing the loose jet identification requirements, there should be no jet passing the loose b jet selection.

Figure 7.3 shows the dimuon p_{T} distributions after the dimuon selections and after the $\text{Z} \to \mu\mu$ selections. The p_{T} of the dimuons from the Z decay usually has very low p_{T} (around 5 GeV), while from the $\text{t}\bar{\text{t}}$ process it tends to have high p_{T}. The Z mass window cut and the b jet veto reduce significantly the $\text{t}\bar{\text{t}}$ contributions in the $\text{Z} \to \mu\mu$ sample, compared with the dimuon sample.

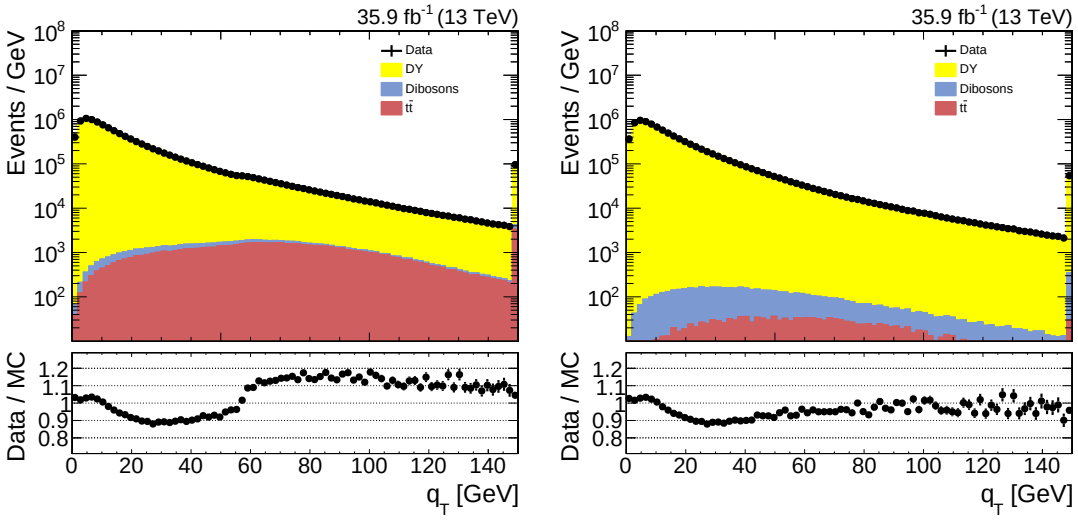


Figure 7.3: Distribution of the p_{T} of the dimuons after the dimuon selection (left) and after the $\text{Z} \to \mu\mu$ selections (right). The underflow (overflow) contents are included in the first (last) bin.

7.2.3 Pileup reweighting

The MC samples are generated with pileup interactions that are meant to roughly cover the conditions expected for 2016 data-taking. However, the actual pileup in data can still be different, depending on beam conditions, instantaneous luminosity, etc. Events in the simulated samples are weighted such that the resulting pileup distribution matches that of data, calculated from the measured instantaneous luminosity and the total inelastic pp cross section $\sigma_{in}^{pp} = 69.2\,\text{mb}$ [40, 41].

The distribution of the number of reconstructed primary vertices (PV) is sensitive to the details of the PV reconstruction and the underlying events in data and MC. It can also be biased by the offline event selection criteria. In order to factorize these effects, instead of reweighting the MC by the number of reconstructed PVs, the number of pileup interactions from the simulation truth is used for the reweighting, and the target pileup distribution for data is derived using the measured instantaneous luminosity.

To validate the reweighting technique, the comparisons of the number of reconstructed primary vertices and the average energy density ρ (which is less affected by non-linear efficiency in the vertex reconstruction) are done between the data and the simulation after reweighting, shown in Figure 7.4 and Figure 7.5. After the reweighting, the MC can reproduce the pileup distributions observed in data, within the estimated uncertainties.

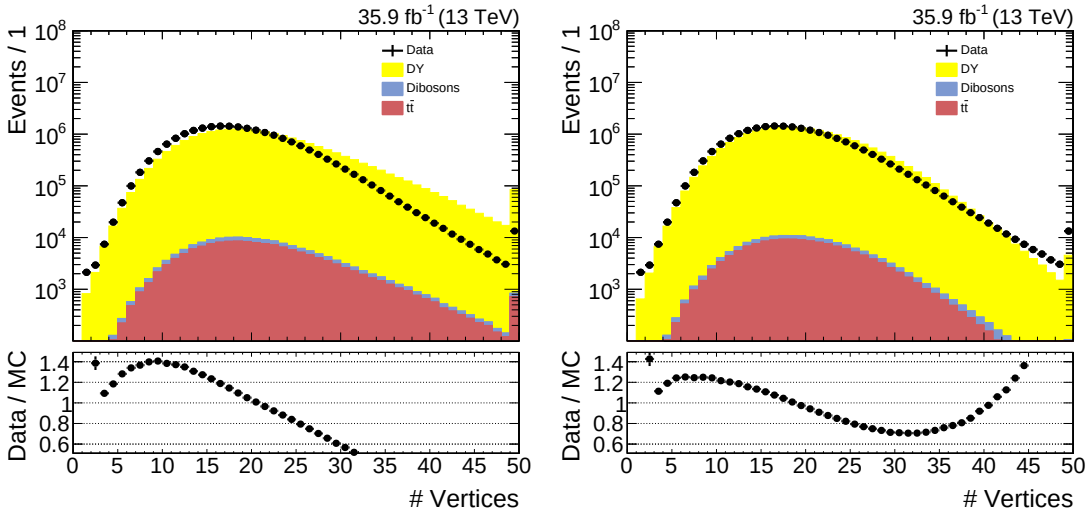


Figure 7.4: Distribution of the number of reconstructed PVs before (left) and after (right) pileup reweighting for events after the dimuon selections. The underflow (overflow) contents are included in the first (last) bin. The data-MC difference is within 20% in the central region after the pileup reweighting.

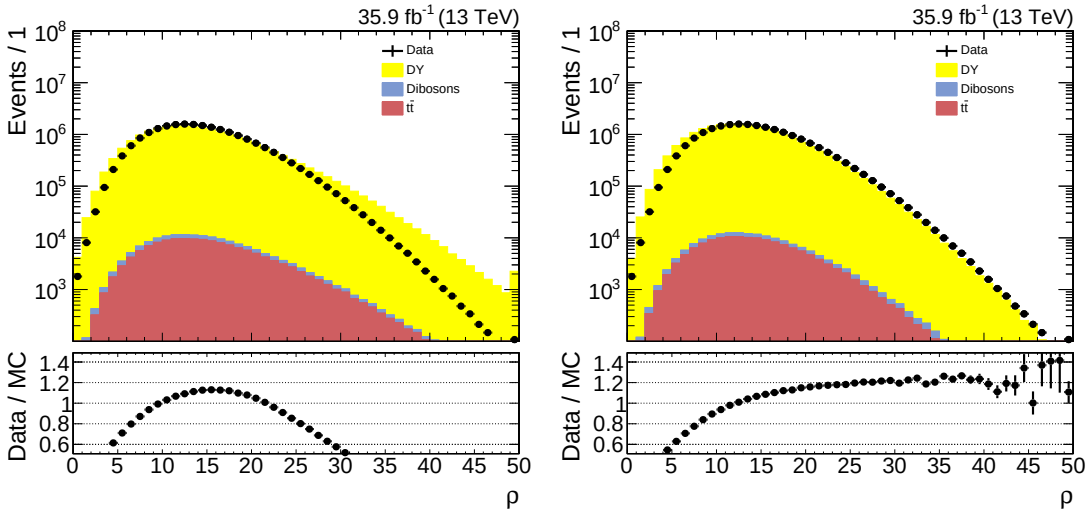


Figure 7.5: Distribution of the average energy density ρ before (left) and after (right) pileup reweighting for events after the dimuon selections. The underflow (overflow) contents are included in the first (last) bin. The data-MC difference is within 20% in the central region after the pileup reweighting.

7.3 DeepMET DNN Setup

The DeepMET DNN setup is discussed in this section, including the inputs and outputs, the architecture, and the loss function used in the training.

7.3.1 Inputs and Outputs

The DNN of DeepMET is trained on the DYJetsToLL and TTJets_DiLept MC samples listed in Table 7.1. Events are required to pass the dimuon selections listed in Section 7.2.2. The purpose of using the two samples is to cover a wide range of p_{T} distributions, as shown in Figure 7.3: the DYJets events tend to have low p_{T} (peak around 5 GeV); the $\mathrm{t\bar{t}}$ events tend to have high p_{T} (peak around 60 GeV with a wide spread). It is also shown in the studies that the DeepMET performance is not very sensitive to different mixture fractions of the two samples. With the current setup, events in data passing the $\mathrm{Z} \to \mu\mu$ selections can also be used in the training, in order to improve the data-MC agreement.

All the PF candidates, except the two leading muons, are passed into the DNN inputs after the event selections. The inputs include 8 continuous features and 3 categorical features for each PF candidate:

- Continuous features:
  - *dxy*: 2D impact parameter in the transverse plane. For neutral PF candidates, the PF reconstruction algorithm always assigns the value of 0;
  - *dz*: The distance between the PF candidate and the LV along the beam

direction. For neutral PF candidates it is always 0;

- η;
- *mass*;
- *PUPPI weights*;
- $p_x = p_{\mathrm{T}} \cdot \cos(\phi)$;
- $p_y = p_{\mathrm{T}} \cdot \sin(\phi)$; here p_x and p_y are used rather that the azimuthal angle ϕ, because sometimes the rotational symmetry is a hard feature for the DNN to learn; and the usage of p_x and p_y also allows the fast calculation of weighted p_{T} sum in the x and y directions;
- p_{T};

- Categorical features:
  - *pdgId*: particle ID assigned to each PF candidates. The possible values are 1, 2, 11, 13, 22, 211. Among these, 11, 13, 22 represent PF electrons, muons and photons, respectively; 211 represents charged PF hadrons; 130 represents neutral PF hadrons; 1 and 2 represent hadronic and electromagnetic PF candidates in the HF;
  - *charge*: ± 1, 0;
  - *fromPV*: the fitting status between the LV and the charged PF candidate. The possible values are 0, 1, 2, 3. Among these, 0 means the candidate is used in the fit of one of the PVs other than the LV; 1 means the candidate is closer in the beam direction to one of the PVs other than the LV; 2

means the candidate is not used in any PV fit, and is closer in the beam direction to the LV than to any other PVs; 3 means the charged PF candidate is used in the fit of the LV. The *fromPV* value for neutral PF candidates is always 3.

The outputs of the DNN are the hadronic recoils u_{T} in the x and y direction: u_x and u_y. The target values for the training are calculated using the vector p_{T} sum of the two reconstructed muons, multiplied by -1 to flip the direction. Considering the resolution differences between the recoil and the muon p_{T} measurements, there is little difference using the reconstructed muon p_{T} and the generator-level muon p_{T} in the training.

7.3.2 DNN Architecture

Figure 7.6 shows the DNN architecture. The inputs are the 11 features of each PF candidate. After the input layer, each of the three categorical features will go through one embedding layer and become an 8-dimensional tensor, such that the feature differences can be represented by distances in the embedded space. Figure 7.7 shows the 2D projection of the embedded particle pdgIds after the training, using a principal component analysis [105, 106]. The distance between the charged hadrons with opposite charges is small, which shows they have similar properties, although their pdgIds are opposite: from -211 to 211; the distance between HF_e and HF_h, electromagnetic and hadronic HF candidates, respectively, is small as well; the distance between charged hadrons, neural hadrons, photons, and HF candidates

are large, indicating that they have different properties, as expected.

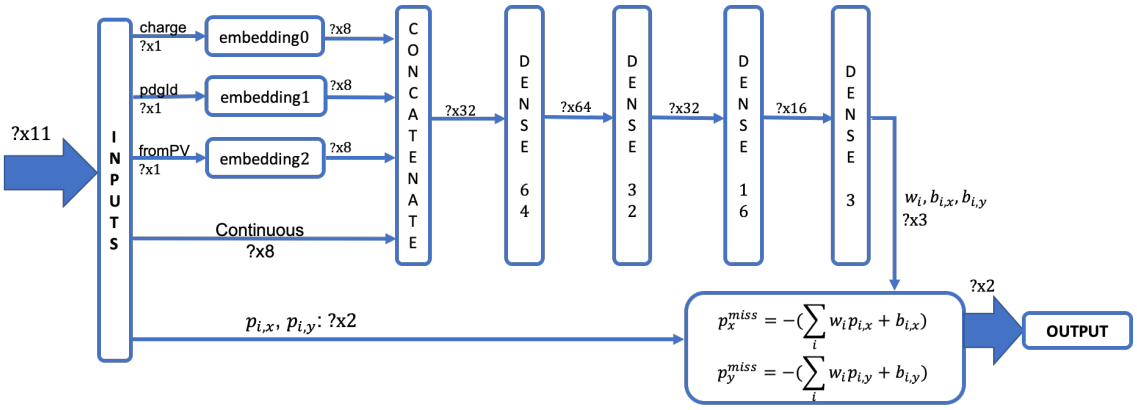


Figure 7.6: DeepMET DNN architecture.

The three 8-dimensional tensors from the embedding layers are then concatenated together with the 8 continuous features, and passed to a few fully connected dense layers. After each dense layer, Batch Normalization [93] is applied (not shown in Figure 7.6 for simplicity) to reduce the covariate shifts and make the training converge faster. The outputs of these dense layers are the weights w_i, and the biases $b_{i,x}$ and $b_{i,y}$ of the PF candidates. In the last layer, the event recoil u_{T} will be calculated as follows:

$$u_x = \sum_i (w_i p_{i,x} + b_{i,x}) \,, \tag{7.6}$$

$$u_y = \sum_i (w_i p_{i,y} + b_{i,y}) \,. \tag{7.7}$$

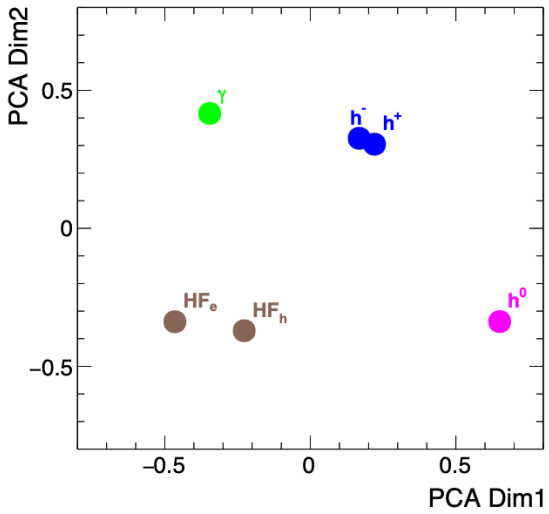


Figure 7.7: Visualization of the embedded particle pdgIds in the 2D space, using a principal component analysis. $h^{\pm}$ represents the charged PF hadrons (pdgId ± 211); γ represents photons (pdgId 22); h^0 indicates neutral PF hadrons (pdgId 130); HF_e and HF_h identify electromagnetic and hadronic HF candidates (pdgId 2 and 1).

The event $p_{\mathrm{T}}^{\mathrm{miss}}$ is then calculated with Equation 7.3, which becomes:

$$p_x^{\mathrm{miss}} = -(u_x + p_{\mathrm{T},x}^{\mu\mu}) \,, \tag{7.8}$$

$$p_y^{\mathrm{miss}} = -(u_y + p_{\mathrm{T},y}^{\mu\mu}) \,. \tag{7.9}$$

This DNN architecture is implemented with Keras [107], which acts as an interface for the Tensorflow [108] library. The total number of trainable parameters in DeepMET is relatively small: only about 5000. It allows the training to converge very fast, and also saves a lot of inference time.

The training is performed on an NVIDIA Quadro P6000 Graphics card, with a few tens of thousand events from the DYJetsToLL sample, and approximately the same number of events from the TTJets_DiLept sample. The batch size is set to 128, and the training and validation splitting is set to 4:1. Further studies are done

by scanning the hyperparameters, such as the mixture fractions of the two samples, the training and validation splitting, and the batch size, and by increasing the number of total events in the training and validation samples. It is observed that the DeepMET performance is weakly sensitive to modifications of the hyperparameters within reasonable ranges. With this setup, the training takes about 30 minutes to converge to stable results.

7.3.3 Loss Function

The loss function used in the training is:

$$
\begin{aligned}
L = \quad & \frac{1}{2}[(u_x - \hat{u}_x)^2 + (u_y - \hat{u}_y)^2] \\
& + c \cdot [(\sum_{\mathrm{R}>1} 1 - \mathrm{R}) - (\sum_{\mathrm{R}<1} \mathrm{R} - 1)] \cdot (q_{\mathrm{T}} < 5) \\
& + c \cdot [(\sum_{\mathrm{R}>1} 1 - \mathrm{R}) - (\sum_{\mathrm{R}<1} \mathrm{R} - 1)] \cdot (q_{\mathrm{T}} > 5 \&\& q_{\mathrm{T}} < 10) \\
& + c \cdot [(\sum_{\mathrm{R}>1} 1 - \mathrm{R}) - (\sum_{\mathrm{R}<1} \mathrm{R} - 1)] \cdot (q_{\mathrm{T}} > 10)\,.
\end{aligned}
\tag{7.10}
$$

where $u_{x,y}$, $\hat{u}_{x,y}$ are the predicted and truth recoils in the x, y direction, R is the response $-\langle u_{\parallel} \rangle / \langle q_{\mathrm{T}} \rangle$, and c is a hyperparameter.

The first part of the loss function is the standard mean square error (MSE) loss function, which is often used in regression tasks. It has been tested and found that, with only the MSE loss, the DeepMET responses saturate around 90% in the high q_{T} region and can not achieve a value of unity. Therefore the second part is added, with the purpose of penalizing the average responses far from unity by balancing

the response above 1 and below 1 in different q_{T} bins. The response is significantly improved with the full loss function. It flattens around unity at high q_{T}, which will be shown in the next section.

The parameter c is the weight between the MSE loss and the response balance term. A few different values ($c = 2$, $c = 5$, $c = 10$, $c = 200$) have been tested, and similar performances have been observed. In the following, c is set to 10 by default.

7.4 DeepMET Performance

The DeepMET performance in data and MC is discussed in this section. Improved resolutions of $u_{\parallel}$ and $u_{\perp}$ have been observed, with unitary responses and more pileup resilience, compared with the commonly used PF and PUPPI $p_{\mathrm{T}}^{\mathrm{miss}}$.

The PF and PUPPI $p_{\mathrm{T}}^{\mathrm{miss}}$ in this section refer to the PF and PUPPI $p_{\mathrm{T}}^{\mathrm{miss}}$ with the Type-I corrections.

7.4.1 Performance on $Z \to \mu\mu$ MC

Figure 7.8 shows the responses of different recoil estimators as a function of q_{T} in $Z \to \mu\mu$ MC, after the dimuon selections. All three estimators reach approximately unitary response ($-\langle u_{\parallel}\rangle/\langle q_{\mathrm{T}}\rangle$) at high q_{T}, while PUPPI $p_{\mathrm{T}}^{\mathrm{miss}}$ has slightly lower response, especially at low q_{T}. It is also worth noting that, at low q_{T}, the DeepMET response is 5-10% lower than PF $p_{\mathrm{T}}^{\mathrm{miss}}$. This is probably due to the bias-resolution tradeoff and indicates the space for improving the DeepMET response.

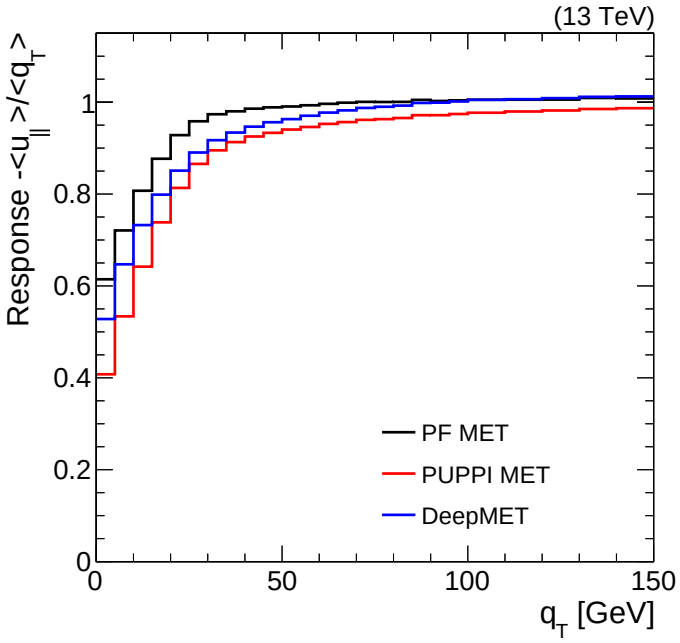


Figure 7.8: Recoil responses of different $p_{\mathrm{T}}^{\mathrm{miss}}$ estimators in $Z \to \mu\mu$ MC after the dimuon event selections.

Table 7.2: Resolutions of $u_{\parallel}$ and $u_{\perp}$ of PF $p_{\mathrm{T}}^{\mathrm{miss}}$, PUPPI $p_{\mathrm{T}}^{\mathrm{miss}}$ and DeepMET with and without response corrections in Z $\to \mu\mu$ MC, after the dimuon selections.

Type	$\sigma(u_{\parallel})$ [GeV]	Corrected $\sigma(u_{\parallel})$ [GeV]	$\sigma(u_{\perp})$ [GeV]	Corrected $\sigma(u_{\perp})$ [GeV]
PF $p_{\mathrm{T}}^{\mathrm{miss}}$	21.1	25.2	20.6	24.7
PUPPI $p_{\mathrm{T}}^{\mathrm{miss}}$	15.4	21.9	14.1	20.1
DeepMET	13.2	17.1	11.9	15.5

Figure 7.9 and 7.10 show the resolutions (defined as $\sigma(q_{84} - q_{16})/2$, as in Equation 7.5) of $u_{\parallel}$ and $u_{\perp}$ *vs.* q_{T} before and after the response corrections, for different recoil estimators in $Z \to \mu\mu$ MC after the dimuon selections. The response corrections are applied by scaling the recoil estimations with the reciprocal of their responses (binned in q_{T}) and measuring the resolutions of the scaled recoil estimations. DeepMET always has the smallest resolutions among all the $p_{\mathrm{T}}^{\mathrm{miss}}$ estimators under consideration in these distributions. The numerical resolution comparisons of $u_{\parallel}$ and $u_{\perp}$ with and without response corrections are reported in Table 7.2.

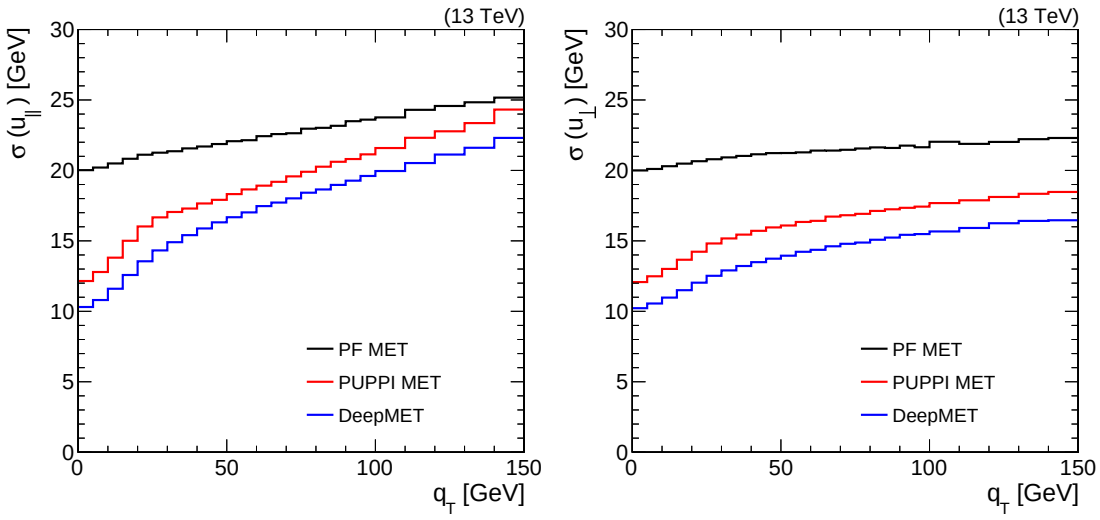


Figure 7.9: Resolutions of $u_{\parallel}$ (left) and $u_{\perp}$ (right) *vs.* q_{T} of different $p_{\mathrm{T}}^{\mathrm{miss}}$ estimators in $\mathrm{Z} \to \mu\mu$ MC, after the dimuon selections.

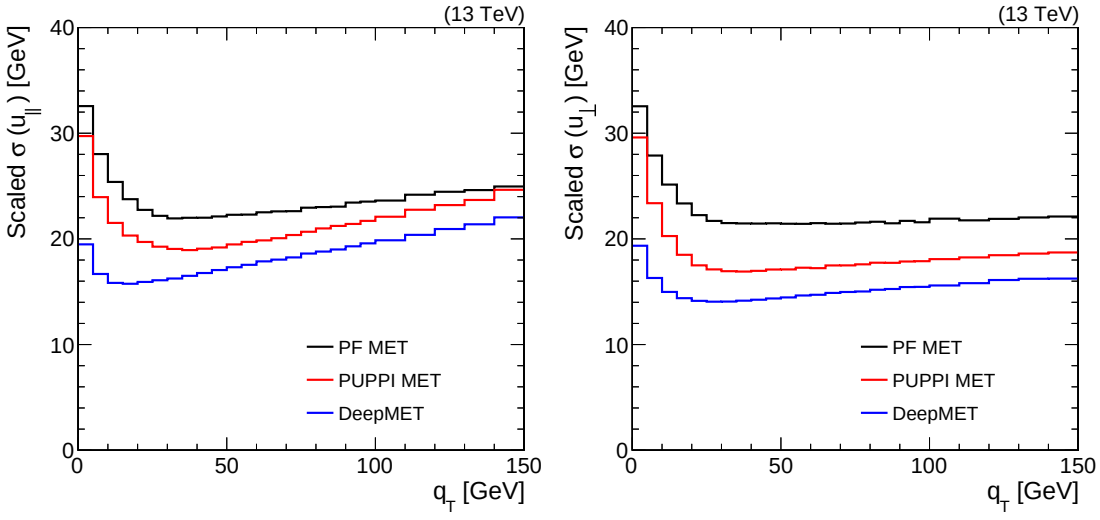


Figure 7.10: Resolutions of $u_{\parallel}$ (left) and $u_{\perp}$ (right) *vs.* q_{T} of different $p_{\mathrm{T}}^{\mathrm{miss}}$ estimators in $\mathrm{Z} \to \mu\mu$ MC, after the dimuon selections. The resolutions are corrected with the responses taken from Figure 7.8.

Figure 7.11 shows the response-corrected resolution distributions of $u_{\parallel}$ and $u_{\perp}$, from the same events, but as a function of the number of reconstructed PVs. Besides the overall smallest resolutions, the resolution variations of DeepMET as the number of PVs increases are the smallest as well, proving its resiliency against pileup conditions. At very low pileup (the number of PVs smaller than 5), the resolutions of PUPPI $p_{\mathrm{T}}^{\mathrm{miss}}$ and DeepMET are slightly above PF $p_{\mathrm{T}}^{\mathrm{miss}}$. This is because many particles, especially the neutrals, produced from the LV are down-weighted, leading to relatively worse performances in this specific region. While this will not affect the majority of analyses, it indicates that there is room for improving DeepMET for low pileup events.

The DeepMET performance has also been tested on other MC samples ($\mathrm{Z}(\nu\nu)+$ jets, $\mathrm{W}(\mu\nu)+$ jets, $\gamma+$ jets, etc.), and similar performance and improvements with respect to the PF and PUPPI $p_{\mathrm{T}}^{\mathrm{miss}}$ estimators have been observed.

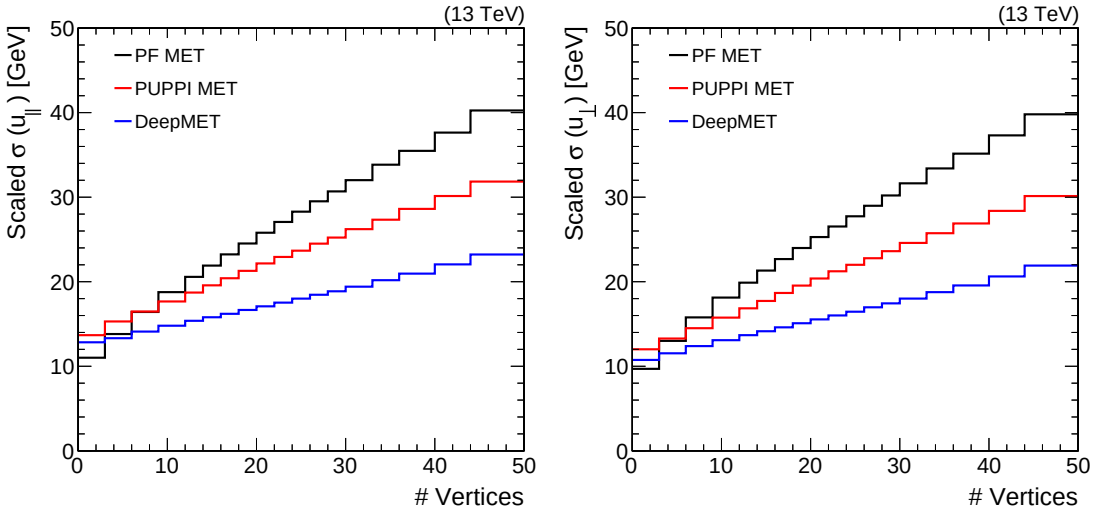


Figure 7.11: Resolutions of $u_{\parallel}$ (left) and $u_{\perp}$ (right) *vs.* number of reconstructed PVs of different $p_{\mathrm{T}}^{\mathrm{miss}}$ estimators in $Z \to \ell\ell$ MC, after the dimuon selections. The resolutions are corrected with the responses taken from Figure 7.8.

Table 7.3: Resolutions of $u_{\parallel}$ and $u_{\perp}$ of PF $p_{\mathrm{T}}^{\mathrm{miss}}$, PUPPI $p_{\mathrm{T}}^{\mathrm{miss}}$ and DeepMET with and without response corrections in data, after the $\mathrm{Z} \to \mu\mu$ selections.

Type	$\sigma(u_{\parallel})$ [GeV]	Corrected $\sigma(u_{\parallel})$ [GeV]	$\sigma(u_{\perp})$ [GeV]	Corrected $\sigma(u_{\perp})$ [GeV]
PF $p_{\mathrm{T}}^{\mathrm{miss}}$	19.8	25.3	19.4	24.8
PUPPI $p_{\mathrm{T}}^{\mathrm{miss}}$	13.5	21.4	12.1	19.6
DeepMET	12.2	17.6	11.1	16.1

7.4.2 Performance on $Z \to \mu\mu$ data

The DeepMET algorithm is applied to the CMS 2016 datasets collected with the SingleMuon trigger, in order to validate the improvements in data. The events are required to pass the $\mathrm{Z} \to \mu\mu$ selections in Section 7.2.2, in order to veto data events with genuine $p_{\mathrm{T}}^{\mathrm{miss}}$ (mostly from electroweak decays inside the heavy flavor jets) and guarantee $\vec{q}_{\mathrm{T}} + \vec{u}_{\mathrm{T}} = 0$, as stated previously.

The responses *vs.* q_{T}, and the resolutions of $u_{\parallel}$ and $u_{\perp}$ *vs.* q_{T} after the event selections are shown in Figure 7.12 and 7.13 respectively. Response-corrected resolutions of $u_{\parallel}$ and $u_{\perp}$ *vs.* q_{T} and *vs.* number of PVs are shown in Figure 7.14 and 7.15. Table 7.3 includes the $u_{\parallel}$ and $u_{\perp}$ resolution values with and without response corrections. In data, DeepMET maintains a unitary response, better resolutions and pileup resilience. Compared with the performances in $\mathrm{Z} \to \mu\mu$ MC, there are some minor differences, which can be eliminated with the quantile corrections that will be discussed in the later sections.

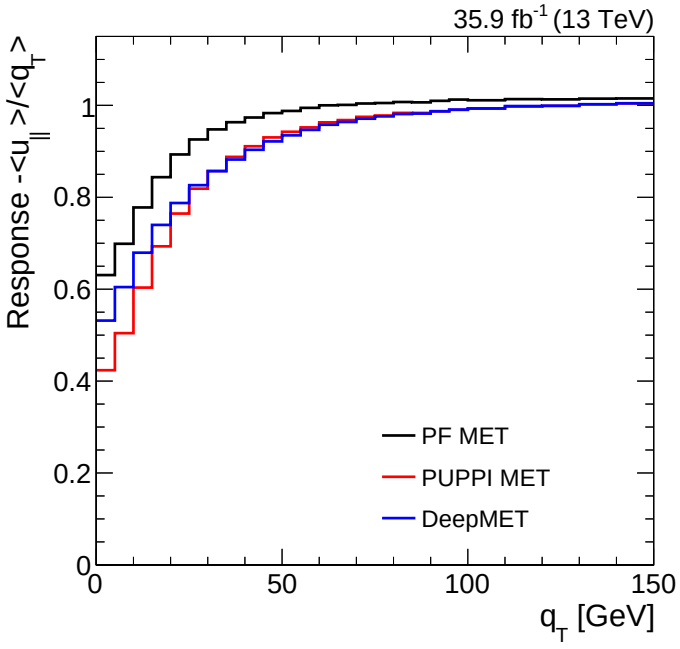


Figure 7.12: Recoil responses of different $p_{\mathrm{T}}^{\mathrm{miss}}$ estimators in data, after the $\mathrm{Z} \to \mu\mu$ selections.

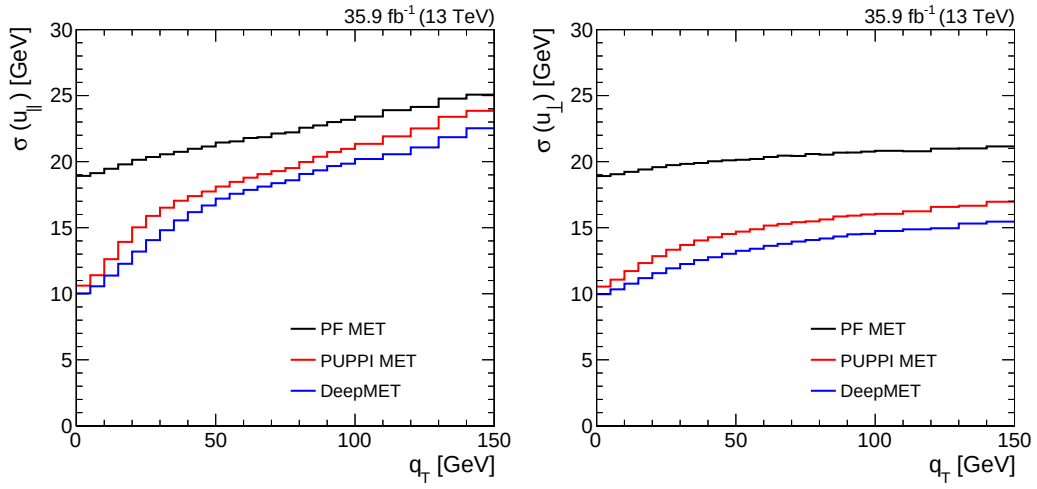


Figure 7.13: Resolutions of $u_{\parallel}$ (left) and $u_{\perp}$ (right) *vs.* q_{T} of different $p_{\mathrm{T}}^{\mathrm{miss}}$ estimators in data, after the $\mathrm{Z} \to \mu\mu$ selections.

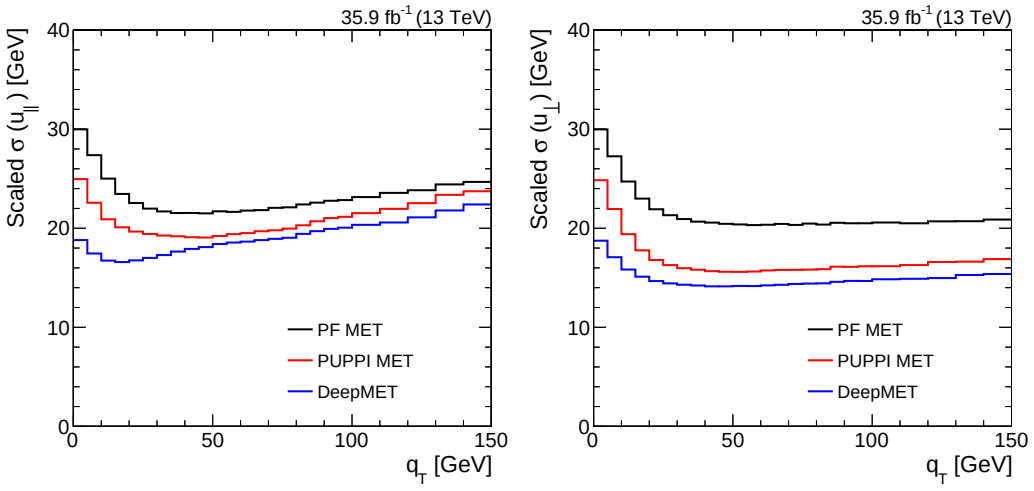


Figure 7.14: Resolutions of $u_{\parallel}$ (left) and $u_{\perp}$ (right) *vs.* q_{T} of different $p_{\mathrm{T}}^{\mathrm{miss}}$ estimators in data, after the $\mathrm{Z} \to \mu\mu$ selections. The resolutions are corrected with the responses taken from Figure 7.12.

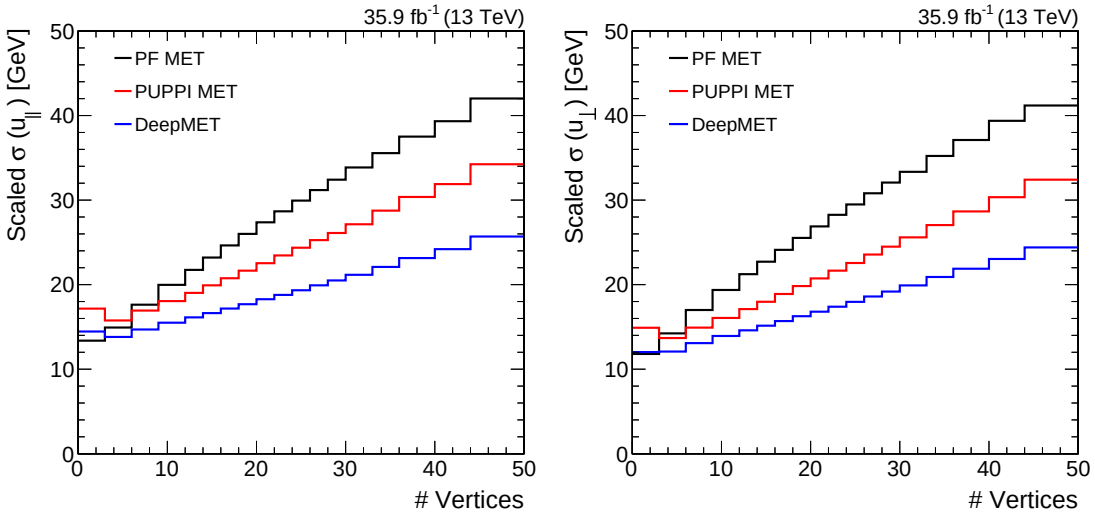


Figure 7.15: Resolutions of $u_{\parallel}$ (left) and $u_{\perp}$ (right) *vs.* number of reconstructed PVs of different $p_{\mathrm{T}}^{\mathrm{miss}}$ estimators in data, after the $\mathrm{Z} \to \mu\mu$ selections. The resolutions are corrected with the responses taken from Figure 7.12.

7.5 A Deeper Look into DeepMET

Further studies have been performed to understand better DeepMET and the origin of its improvements. These additional studies are documented in this section.

7.5.1 Training without PUPPI Weights

It is interesting to check the DeepMET performance in the absence of PUPPI weights information, in order to understand the importance of the PUPPI weights in the inputs. Figure 7.16 shows the response-corrected resolution comparisons of DeepMET trained with PUPPI (solid blue) and without PUPPI weights (dashed blue). The trainings are tested on a small set of MC events so the statistics are limited. Nevertheless, even though the resolutions of DeepMET are worse without PUPPI weights information, they are still systematically smaller than the resolutions of PUPPI and PF $p_{\mathrm{T}}^{\mathrm{miss}}$. This demonstrates that both PUPPI weights and the DeepMET training contribute to the knowledge of $p_{\mathrm{T}}^{\mathrm{miss}}$, and that these two contributions can be combined for further improvements on the $p_{\mathrm{T}}^{\mathrm{miss}}$ estimation.

7.5.2 Contributions from the Bias Terms

The contributions from the bias terms $b_{i,x}$ and $b_{i,y}$ in Equation 7.6 and 7.7 have also been investigated. Figure 7.17 shows the responses and response-corrected resolutions of $u_{\parallel}$ and $u_{\perp}$ *vs.* q_{T}. The solid blue is the original DeepMET, while the dashed blue is the DeepMET removing contributions from the bias terms $b_{i,x}$ and $b_{i,y}$. With almost identical distributions, it can be concluded that the contributions

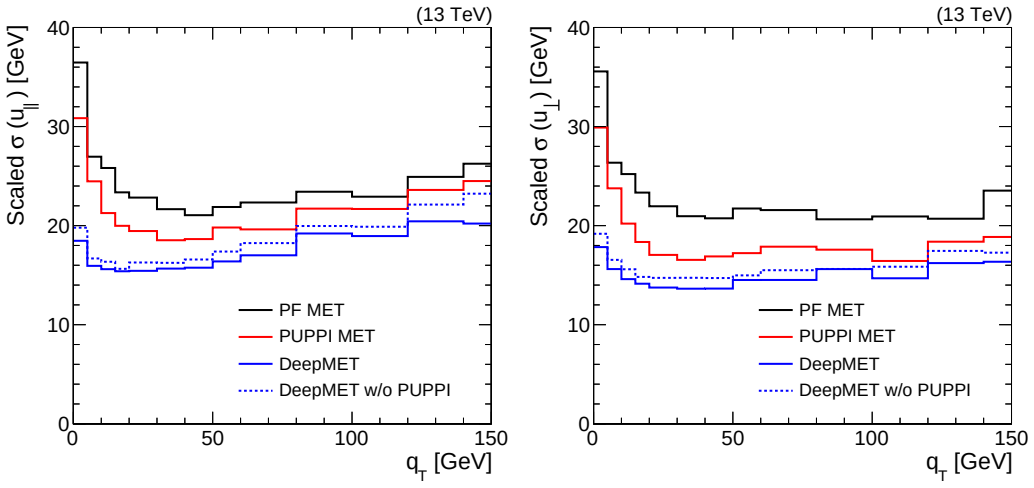


Figure 7.16: Resolutions of $u_{\parallel}$ (left) and $u_{\perp}$ (right) *vs.* q_{T} of different $p_{\mathrm{T}}^{\mathrm{miss}}$ estimators in $\mathrm{Z} \to \mu\mu$ MC, after the dimuon selections. The resolutions are corrected with the responses. The dashed blue distributions are from the DeepMET training without PUPPI weights in the inputs.

from the bias terms are limited, and the main improvements of DeepMET come from the weight term w_i. The bias terms $b_{i,x}$ and $b_{i,y}$ can be used for the corrections in the x-y direction in the future.

7.5.3 DNN Weight per Particle

The distributions of DNN weight per particle w_i are studied in this section. The DNN weights are the key ingredient for the performance improvements since the bias contributions are limited. Figure 7.18 shows the DNN weights for the charged PF hadrons with different *fromPV* flags. As expected, to charged hadrons likely from pileup (*fromPV*$= 0$ and *fromPV*$= 1$), the DNN assigns smaller weights; to charged hadrons likely from the LV (*fromPV*$= 2$ and *fromPV*$= 3$), the DNN assigns higher weights. It is interesting to note that the DNN assigns weights above 1 to charged hadrons from the LV, and weights around 0.2 to charged hadrons from

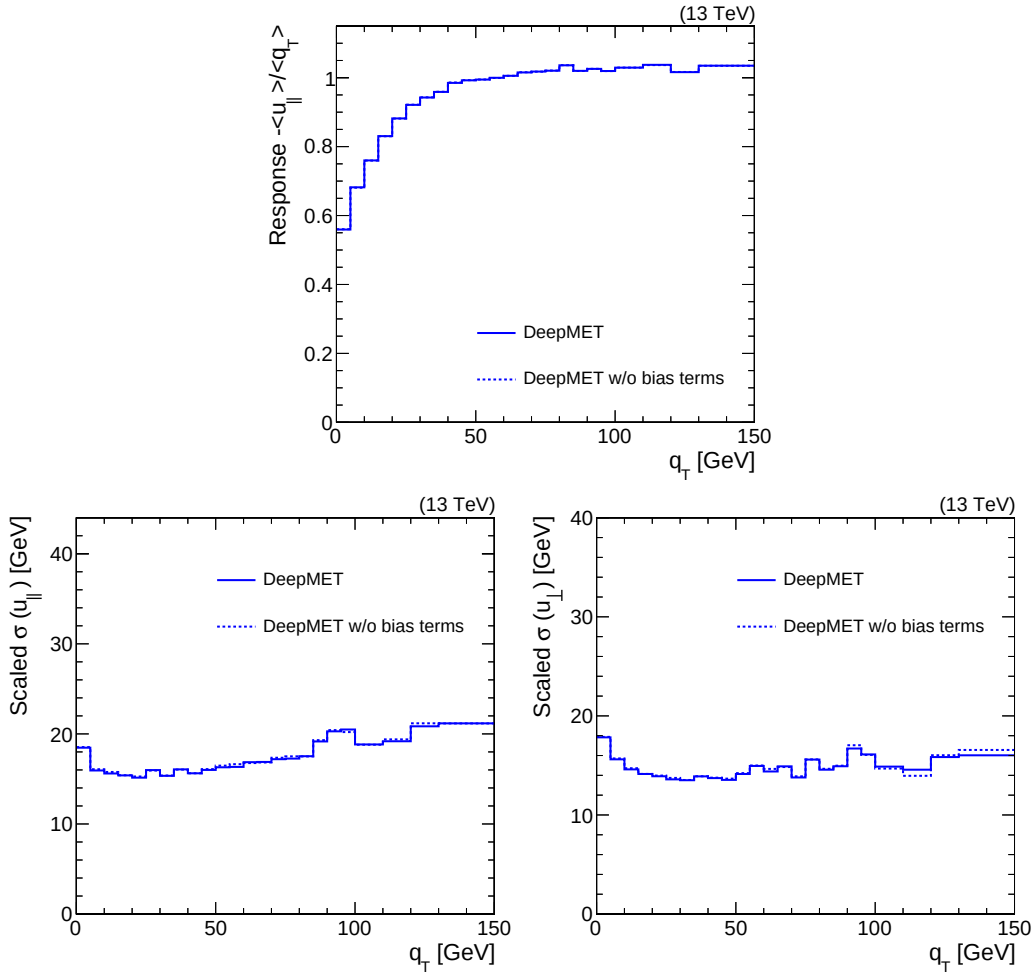


Figure 7.17: Response and resolutions of $u_{\parallel}$ and $u_{\perp}$ of DeepMET with (solid) and without (dashed) contributions from the bias terms, tested in $Z \to \mu\mu$ MC after the dimuon selections. The resolutions are corrected with the responses.

pileup. These behaviors are different from the PUPPI algorithm where charged particles only get assigned binary weight of either 1 or 0.

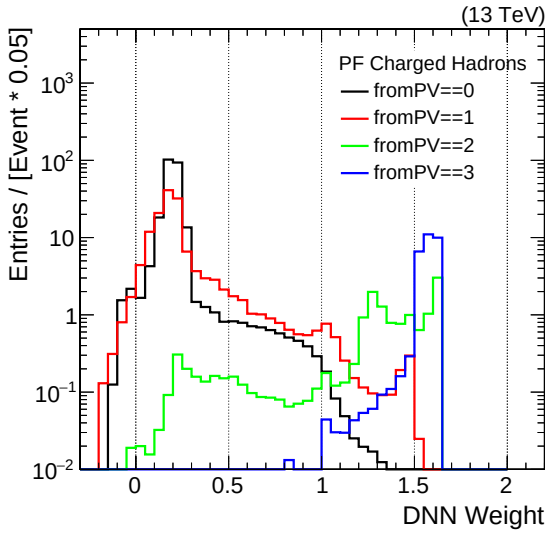
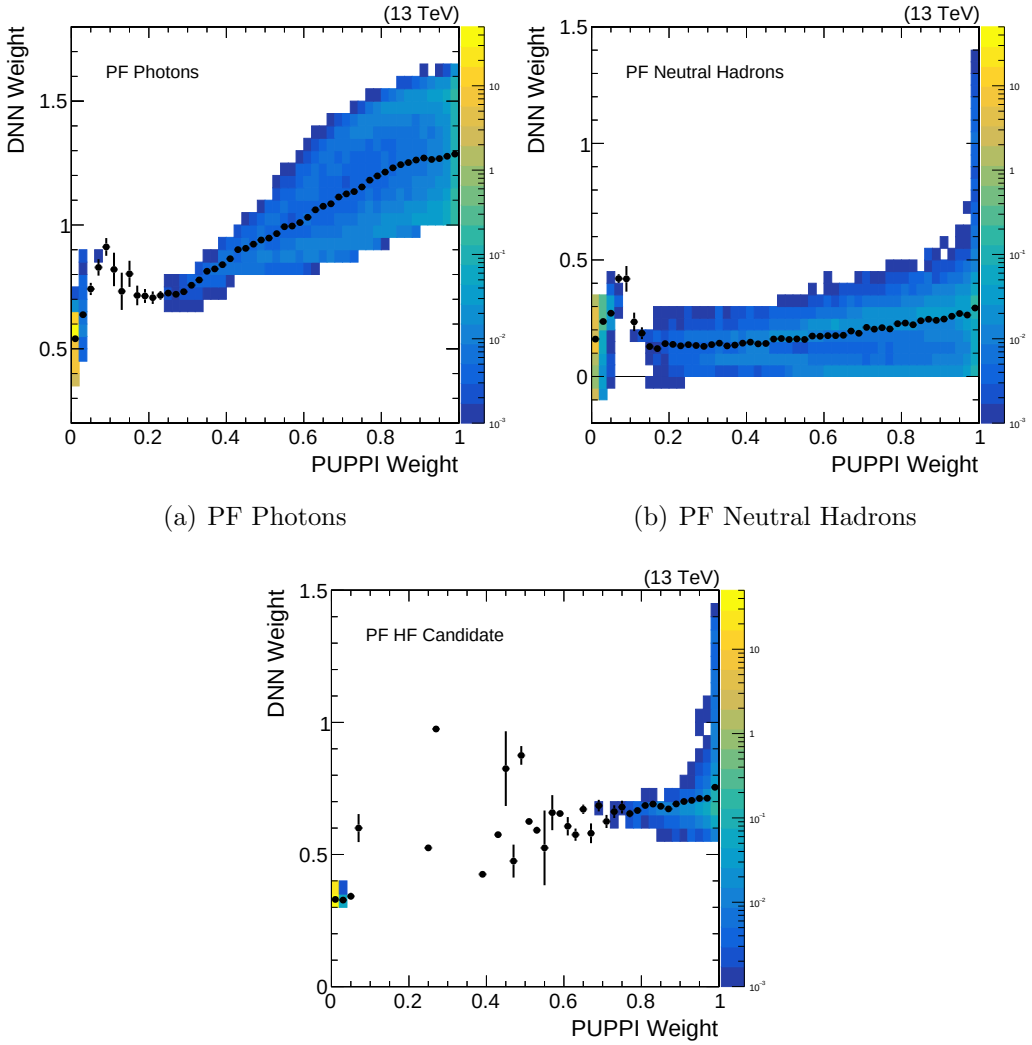


Figure 7.18: DNN weights for charged PF hadrons with different *fromPV* flags.

Figure 7.19 shows the distributions of the DNN weights *vs.* the PUPPI weights for PF photons (top left), PF neutral hadrons (top right) and HF candidates (bottom), and the black distribution in each plot is the profiled DNN weights with respect to PUPPI weights. All distributions have been normalized with the total number of events. For PF photons in Figure 7.19(a), the DNN weights increase as the PUPPI weights increase. For PF neutral hadrons in Figure 7.19(b), the dependence of the DNN weights on the PUPPI weights is much smaller compared with PF photons. The average DNN weights for neutral hadrons is around 0.1 to 0.2, while for PF photons, the weights increase from 0.5 (for the photons with low PUPPI weights) to 1.3 (for the photons with high PUPPI weights).

Figure 7.20 shows the profiled DNN weights *vs.* p_{T} (left) and $|\eta|$ (right) distributions for different types of PF candidates, while Figure 7.21 shows the same sets



(a) PF Photons

(b) PF Neutral Hadrons

(c) HF Candidates

Figure 7.19: DNN weights *vs.* PUPPI weights for PF photons (top left), PF neutral hadrons (top right) and HF candidates (bottom).

for profiled PUPPI weights as comparison. The PF candidates from the LV tend to have higher p_{T}, while the PF candidates from pileup always have low p_{T}. This could explain why the DNN weights increase as p_{T} goes higher. At very high p_{T}, the weights for different types of PF candidates saturate around 1.0. In the weights *vs.* $|\eta|$ distribution, the weights for charged hadrons and HF candidates tend to have little dependence on $|\eta|$, while for PF photons, the weights peak around the gap between ECAL barrel and ECAL endcap ($|\eta| \sim 1.45$). This might be because the reconstruction efficiency for photons is low in that region, and the DNN weights have to increase to compensate for the reconstruction inefficiency.

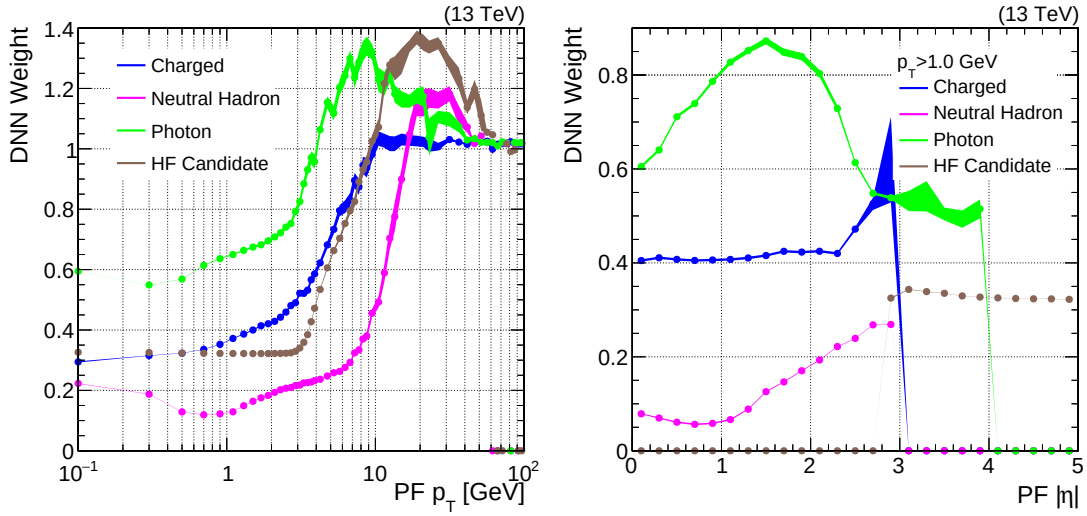


Figure 7.20: Profiled DNN weights *vs.* p_{T} (left) and $|\eta|$ (right) for different PF candidates

7.6 DeepMET Calibrations and Uncertainties

The DeepMET calibrations, with the aim to achieve a better data-MC agreement, are discussed in this section. The uncertainties on DeepMET derived from

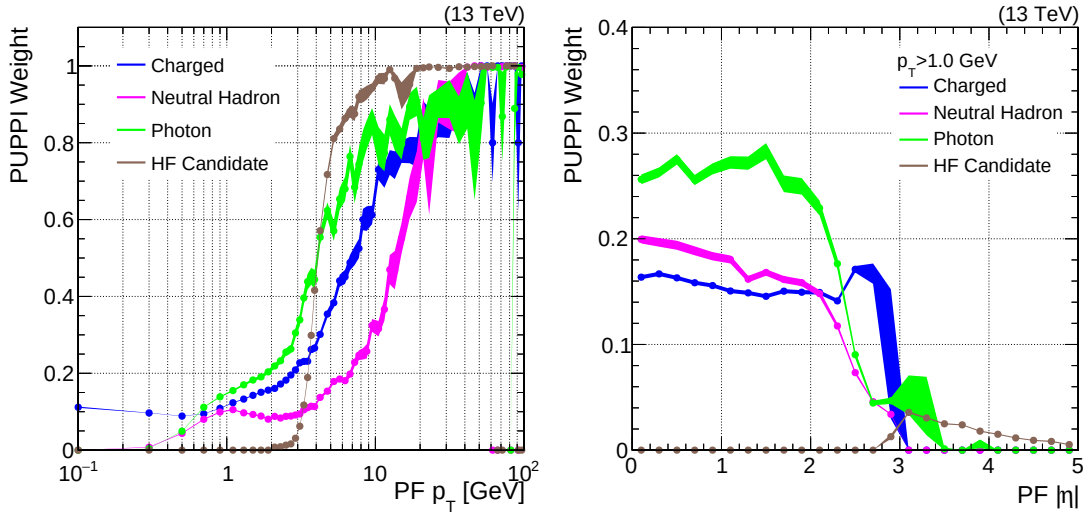


Figure 7.21: Profiled PUPPI weights *vs.* p_{T} (left) and $|\eta|$ (right) for different PF candidates

this calibration process are also defined and estimated.

7.6.1 DeepMET Calibrations

The Z p_{T} reweighting is first applied to the simulation in order to correct the boson p_{T} distributions. The left plot in Figure 7.22 shows the p_{T} distribution of the dimuon pairs before reweighting. The data-MC ratio in the bottom panel is used to reweight the Drell-Yan MC events, as a function of q_{T}. The reweighted distribution is shown in the right plot of Figure 7.22. The MC prediction overestimates the observations in data by about 10% in the high-q_{T} region before the correction. The data-MC ratio becomes flat across the whole spectra after the reweighting.

Figure 7.23 shows the data-MC comparisons of DeepMET (top left), the recoil u_{T} (top right), $u_{\parallel}$ (bottom left) and $u_{\perp}$ (bottom right) distributions before the correction, where the resolution differences in data and MC cause some disagreement between data and MC. A quantile correction is derived and applied as described

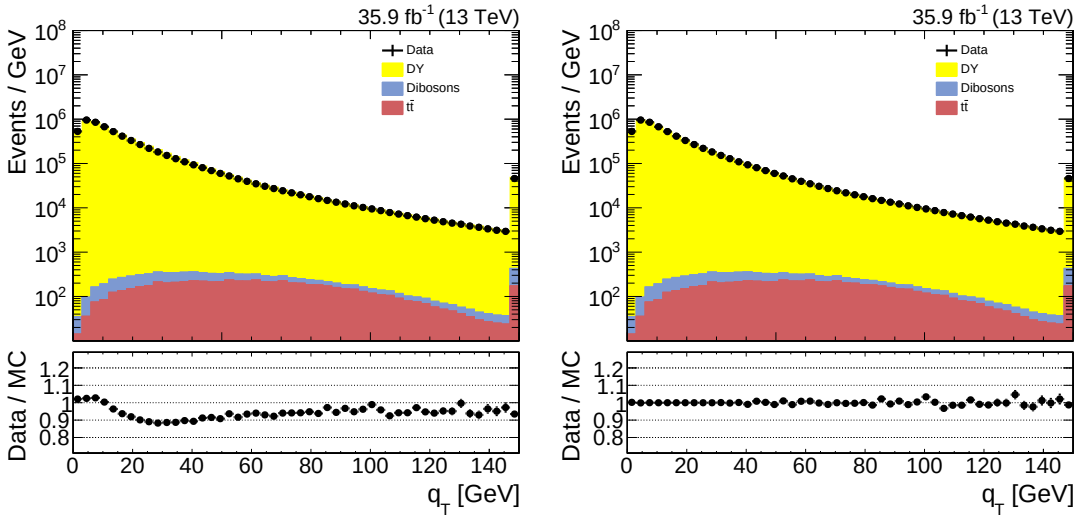


Figure 7.22: Data-MC comparisons of the dimuons p_{T} before (left) and after (right) Z p_{T} reweighting. The underflow (overflow) contents are included in the first (last) bin.

below, in order to correct the distributions and improve the agreement.

First the data and MC are binned in jet multiplicity and Z p_{T}: three bins for jet multiplicity: njet=0, njet=1, and njet>1; 80 bins for Z p_{T} from 0 to 300 GeV. The distributions of $u_{\parallel}$ and $u_{\perp}$ are fitted with double-Gaussian functions in each bin. In data, the $\mathrm{t\bar{t}}$ and diboson contributions have been subtracted using MC templates before the fits. Some examples of the fitting results are shown in Figure 7.24 and 7.25 for data, Figure 7.26 and 7.27 for MC. The distributions are modeled reasonably well by double-Gaussian distributions.

In the second step, each of the fitted double-Gaussian functions is integrated to get the corresponding cumulative distribution functions (CDFs) for data and MC, in different jet multiplicity and Z p_{T} bins. For each $u_{\parallel}$ and $u_{\perp}$ in MC, the CDF

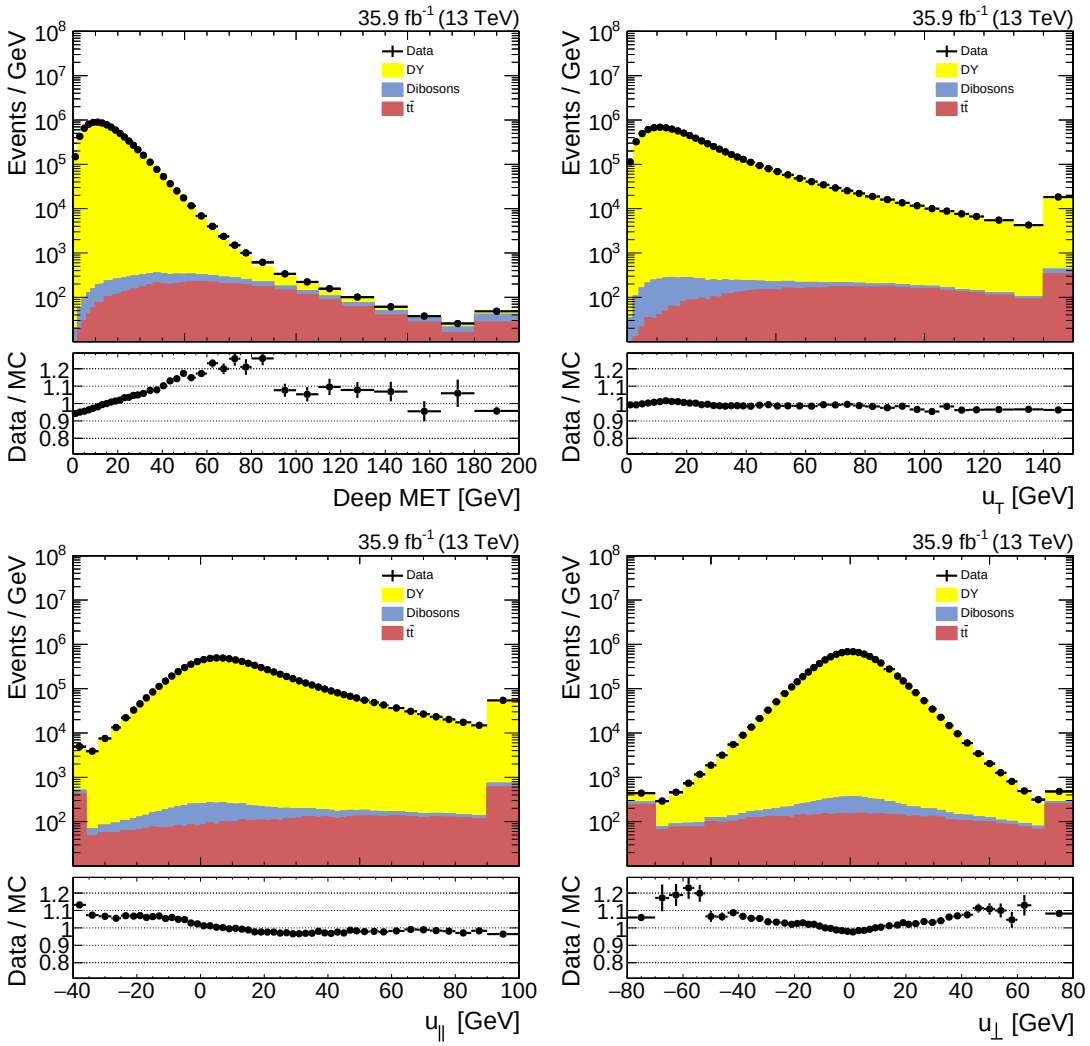


Figure 7.23: Data-MC comparisons of DeepMET p_{T} (top left), the recoil p_{T} (top right), $u_{\parallel}$ (bottom left), and $u_{\perp}$ (bottom right) before the corrections. The underflow (overflow) contents are included in the first (last) bin.

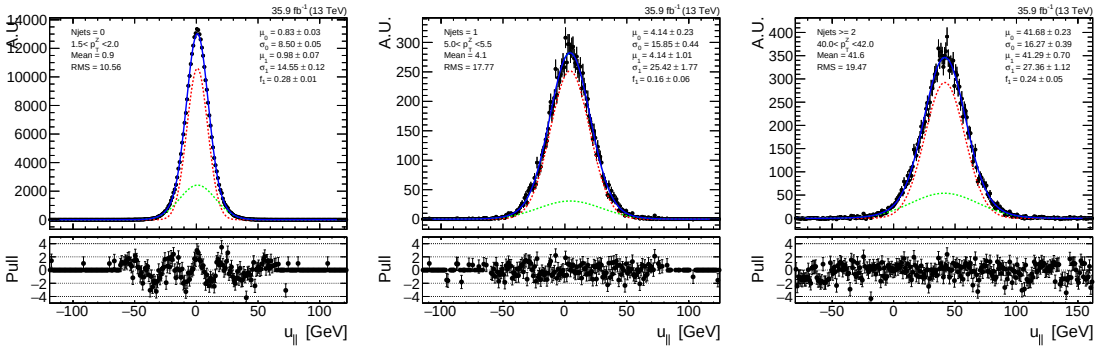


Figure 7.24: Examples of double-Gaussian fits of $u_{\parallel}$ in data, in different jet multiplicity and Z p_{T} bins. The $\mathrm{t\bar{t}}$ and diboson contributions have been subtracted from data using MC templates before the fits.

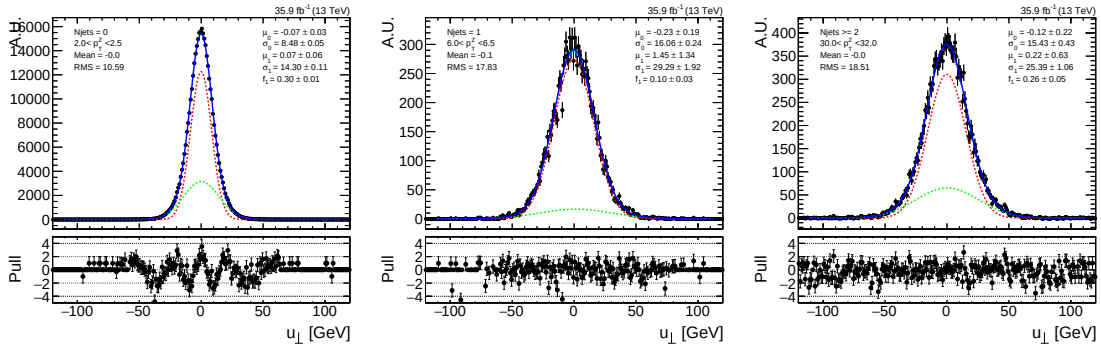


Figure 7.25: Examples of double-Gaussian fits of $u_\perp$ in data, in different jet multiplicity and Z p_T bins. The $\mathrm{t\bar{t}}$ and diboson contributions have been subtracted from data using MC templates before the fits.

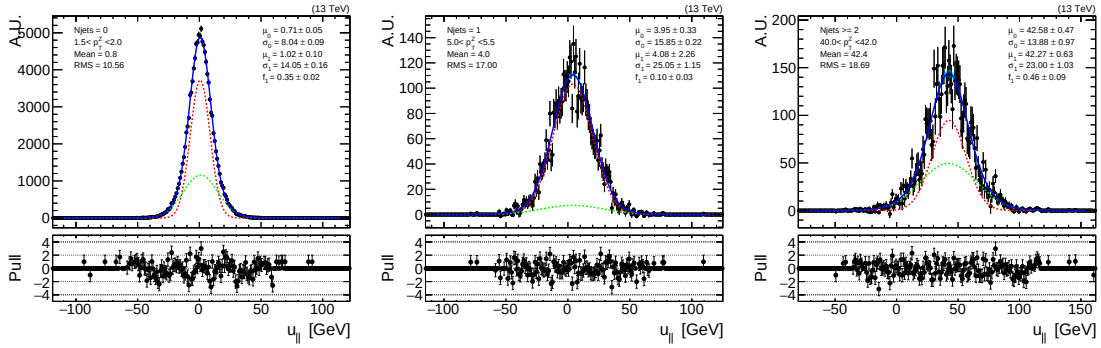


Figure 7.26: Examples of double-Gaussian fits of $u_\parallel$ in Z+jets MC, in different jet multiplicity and Z p_T bins.

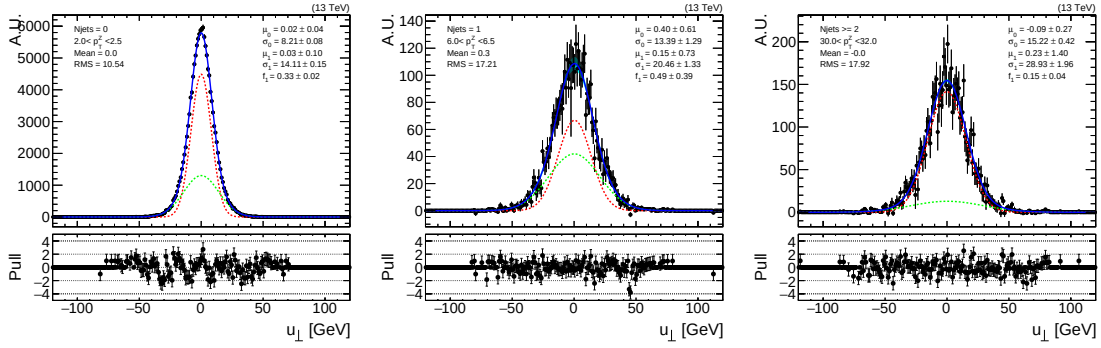


Figure 7.27: Examples of double-Gaussian fits of $u_\perp$ in Z+jets MC, in different jet multiplicity and Z p_T bins.

values can be found from:

$$
\begin{aligned}
p_{\parallel} &= CDF_{i\text{-}bin}^{u_{\parallel},MC}(u_{\parallel}^{MC})\,, \\
p_{\perp} &= CDF_{i\text{-}bin}^{u_{\perp},MC}(u_{\perp}^{MC})\,.
\end{aligned}
\tag{7.11}
$$

In the third step, the corresponding $u_{\parallel}$ and $u_{\perp}$ in data can be found from:

$$
\begin{aligned}
u_{\parallel}^{Data} &= (CDF_{i\text{-}bin}^{u_{\parallel},Data})^{-1}(p_{\parallel})\,, \\
u_{\perp}^{Data} &= (CDF_{i\text{-}bin}^{u_{\perp},Data})^{-1}(p_{\perp})\,.
\end{aligned}
\tag{7.12}
$$

Finally the differences $u_{\parallel}^{Data} - u_{\parallel}^{MC}$ and $u_{\perp}^{Data} - u_{\perp}^{MC}$ are applied as the corrections to $u_{\parallel}$ and $u_{\perp}$ in MC, respectively. The corrected DeepMET is then calculated from the corrected $u_{\parallel}$ and $u_{\perp}$.

Figure 7.28 shows the DeepMET (top left), the recoil p_{T} (top right), $u_{\parallel}$ (bottom left) and $u_{\perp}$ (bottom right) data-MC comparisons after the corrections; good agreements between data and MC are achieved after the quantile correction, with the differences within 10% across the whole spectra.

7.6.2 DeepMET Uncertainties

The uncertainties on DeepMET are defined and calculated from the calibration process. Each of them is listed and discussed below.

1. Uncertainty from MC generators.

   The different corrections derived from simulated events with different generators are compared here. Instead of the Z NLO MC [95], the MadGraph LO

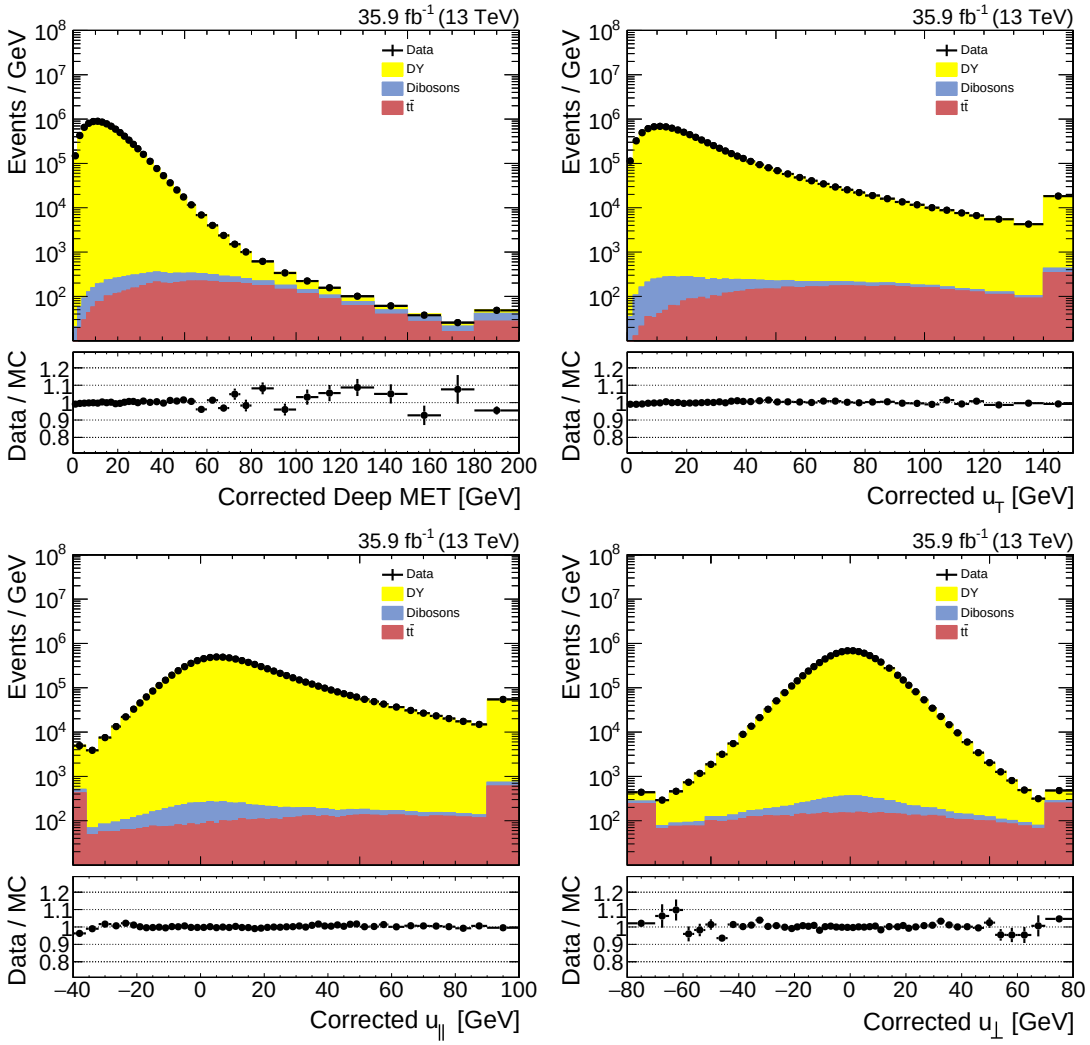


Figure 7.28: Data-MC comparisons of DeepMET p_{T} (top left), the recoil p_{T} (top right), $u_{\parallel}$ (bottom left), and $u_{\perp}$ (bottom right) after the quantile correction. The underflow (overflow) contents are included in the first (last) bin.

sample listed in Table 7.1 is used and the correction process described in the previous section is repeated. The corrections derived from LO sample are applied to the NLO sample. The differences between the two sets of corrections are taken as estimators of the generator uncertainty.

Figure 7.29 shows the comparisons of corrections on $u_{\parallel}$ (top left), $u_{\perp}$ (top right) and DeepMET (bottom). The difference in $u_{\parallel}$ between the two sets of corrections is below a few percents in the whole spectrum; the difference for $u_{\perp}$ in the central region is about 1%, and it ramps up to about 10%, due to resolution effects, at the two edges; the difference for DeepMET is around 1% at low $p_{\mathrm{T}}^{\mathrm{miss}}$, and becomes around 5% in the range above 50 GeV.

2. Uncertainty from the fitting model.

   A double-Gaussian function is used to fit the $u_{\parallel}$ and $u_{\perp}$ distributions in bins of jet multiplicity and Z p_{T}. Instead of a fit, a Gaussian kernel smoothing algorithm [109] is used, with a width parameter set to 1 GeV, to smooth the $u_{\parallel}$ and $u_{\perp}$ distributions in data and MC (similar to a convolution with a Gaussian function). The corrections are derived based on the smoothed distributions rather than the fitted double-Gaussian functions. The differences are used as the estimators of the fitting-model uncertainty.

   Figure 7.30 shows the comparisons of the corrections on $u_{\parallel}$ (top left), $u_{\perp}$ (top right) and DeepMET (bottom). The differences are relatively small: around 1-2% in most of the regions.

3. Uncertainty from the $\mathrm{t\bar{t}}$ background modeling.

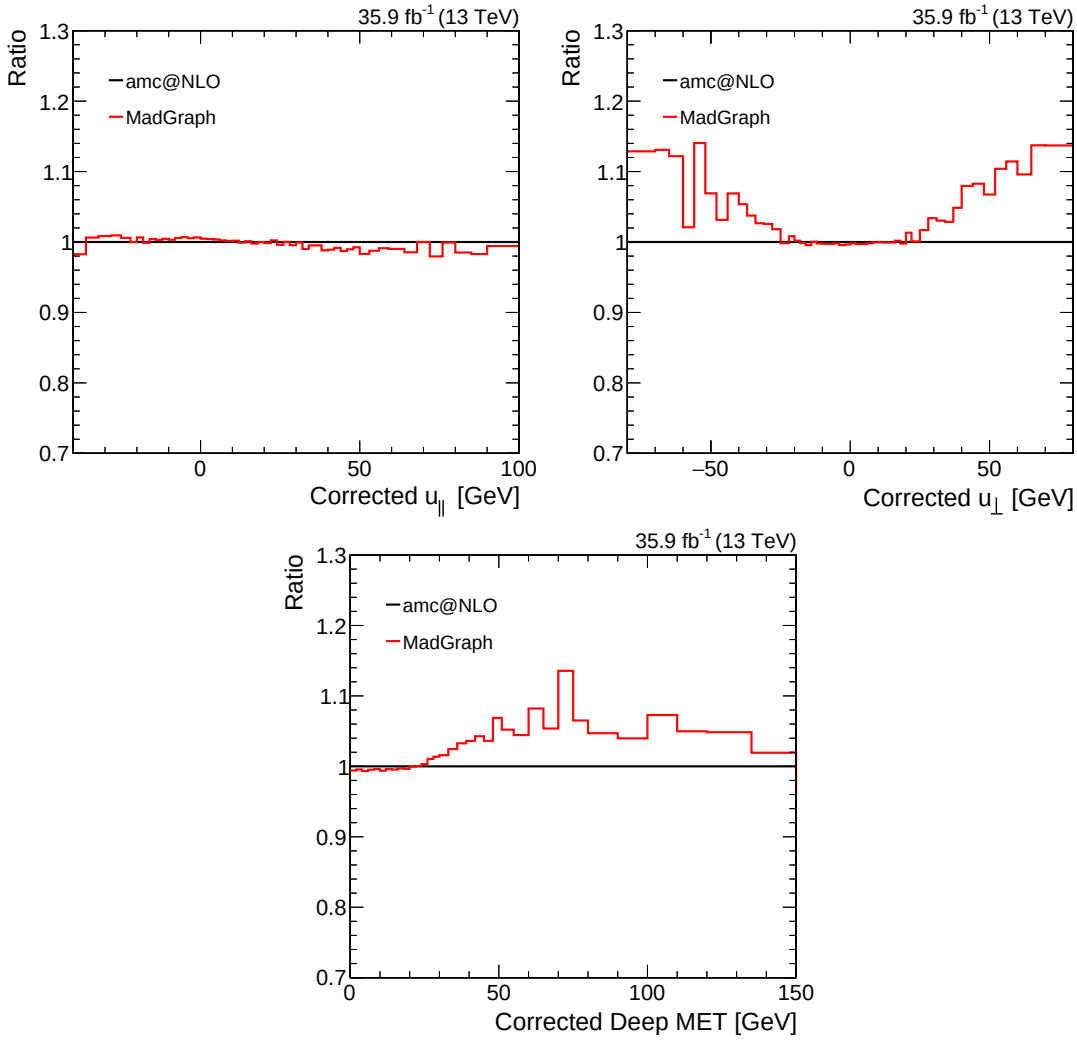


Figure 7.29: Ratio comparison of corrections derived from Z+jets LO over Z+jets NLO MC, on $u_{\parallel}$ (top left), $u_{\perp}$ (top right), and DeepMET (bottom).

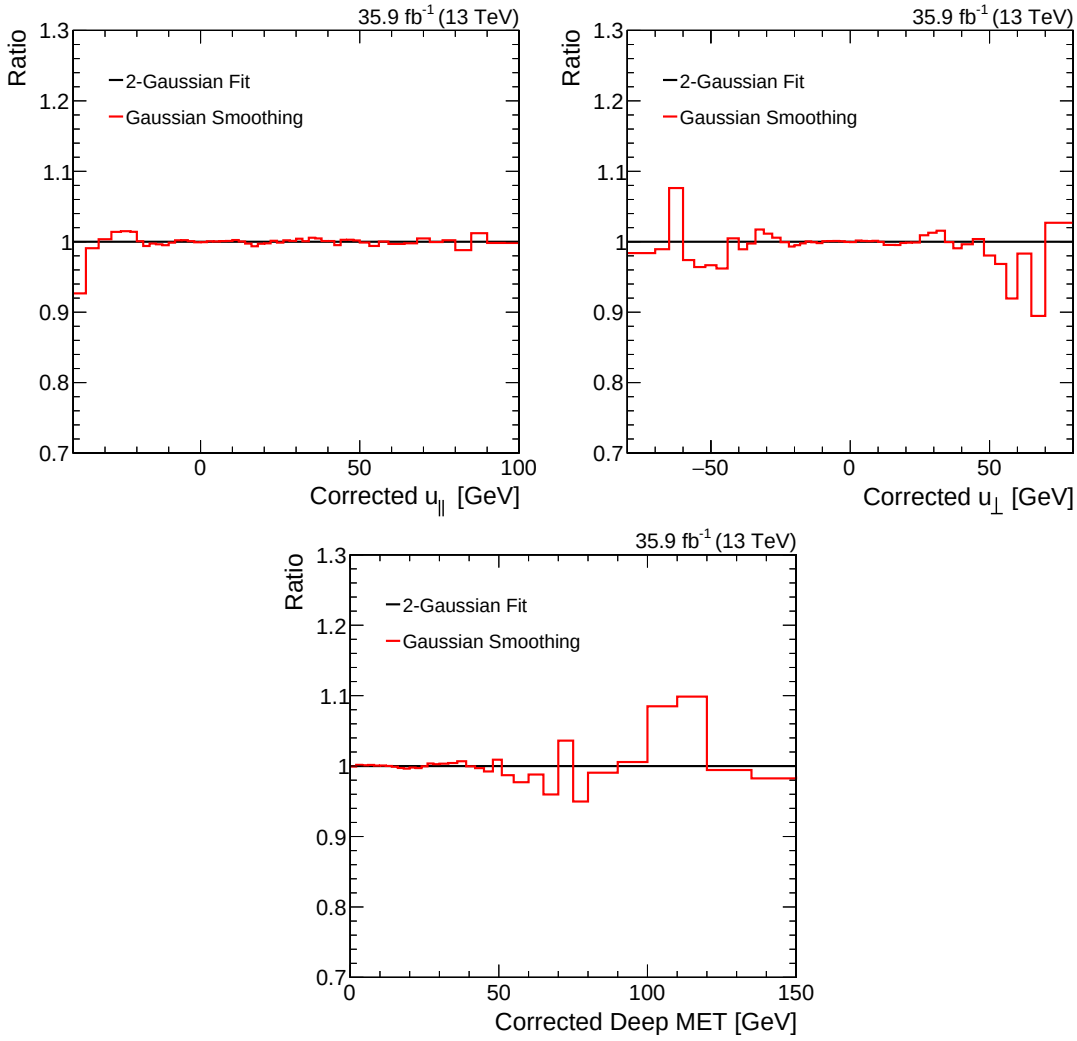


Figure 7.30: Ratio comparison of corrections derived from Gaussian kernel smoothings over double-Gaussian fittings, on $u_{\parallel}$ (top left), $u_{\perp}$ (top right), and DeepMET (bottom).

The $t\bar{t}$ and diboson backgrounds are subtracted using MC templates before the data fitting, where the $t\bar{t}$ background shapes could change as a function of the QCD renormalization and factorization scales. The $u_{\parallel}$ and $u_{\perp}$ templates of the $t\bar{t}$ background are obtained by varying the QCD scales in the simulation. The modified background template are re-subtracted from data and the corrections are re-derived from the new distributions. The differences are the estimators of the systematic uncertainty from the $t\bar{t}$ background modeling.

Figure 7.31 shows the comparisons of the corrections on $u_{\parallel}$ (top left), $u_{\perp}$ (top right) and DeepMET (bottom). The $t\bar{t}$ background mainly contributes to large $u_{\parallel}$ and $u_{\perp}$, as can be seen from Figure 7.28. This explains the distributions remain unchanged for small values of $u_{\parallel}$ and $u_{\perp}$, while the variations are 5-10% variations at large $u_{\perp}$ and large DeepMET. The contribution from the $t\bar{t}$ background is around 1% for $u_{\parallel}$ around 100 GeV, and thus the effect on $u_{\parallel}$ is limited.

In summary, all three uncertainties are relatively small (around 1%) in the central region (where $u_{\parallel}$, $u_{\perp}$, and DeepMET are close to 0); the uncertainties become large (around 5-10%) at large $u_{\parallel}$, $u_{\perp}$ and $p_{\mathrm{T}}^{\mathrm{miss}}$. The uncertainty from MC generators is the dominant uncertainty.

DeepMET is the first DNN-based $p_{\mathrm{T}}^{\mathrm{miss}}$ estimator. Compared with the current PF and PUPPI $p_{\mathrm{T}}^{\mathrm{miss}}$ estimators, it improves the $p_{\mathrm{T}}^{\mathrm{miss}}$ resolution by 10-20% and is more resilient towards pileup, by assigning individual weights to input PF candidates with only about 5000 trainable parameters. Good data-MC agreement is achieved

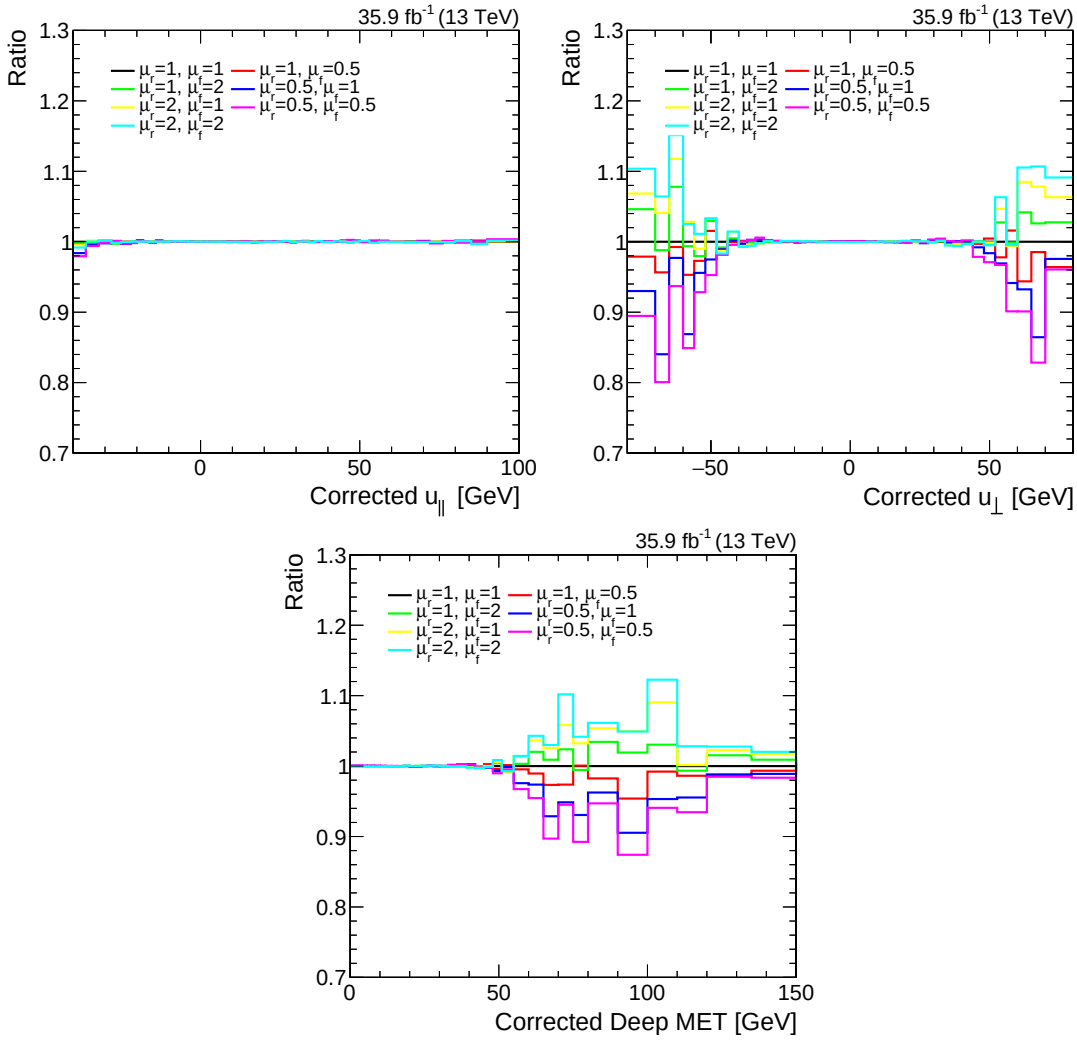


Figure 7.31: Ratio comparison of corrections derived from QCD scale variations of the $t\bar{t}$ background templates, on $u_{\parallel}$ (top left), $u_{\perp}$ (top right), and DeepMET (bottom).

after the quantile correction. DeepMET can help improve the precision of SM measurements and the sensitivity of searches beyond the SM, for processes where $p_{\mathrm{T}}^{\mathrm{miss}}$ is present.

Chapter 8: Study of the Hadronic Recoil Against the W Boson

A precise measurement of the W boson mass (m_{W}) can be used to probe accurately the predictions of the SM and search for indications of new physics [12]. It was performed previously at the SPS $\mathrm{p\bar{p}}$ collider by the UA1 and UA2 collaborations [110, 111] at $\sqrt{s} = 546\,\mathrm{GeV}$ and $\sqrt{s} = 630\,\mathrm{GeV}$, at the Tevatron $\mathrm{p\bar{p}}$ collider by the CDF and D0 collaborations at $\sqrt{s} = 1.8\,\mathrm{TeV}$ [112–114] and $\sqrt{s} = 1.96\,\mathrm{TeV}$ [115–117], at the LEP e^+e^- collider by the ALEPH, DELPHI, L3, and OPAL collaborations at $\sqrt{s}$ =161-209 GeV [118–121], and at the LHC pp collider by the ATLAS Collaboration at $\sqrt{s} = 7\,\mathrm{TeV}$ [122].

The latest measurement from the ATLAS Collaboration achieves a precision of 19 MeV, by making use of a combined fit of the charged lepton p_{T} and the transverse mass (m_T). In this chapter we study the latter variable and explore the potential of reconstructing it with DeepMET.

8.1 Introduction

8.1.1 Transverse Mass m_{T} and the W Mass

W and Z bosons are produced predominantly through quark-antiquark annihilation in pp collisions. Higher-order processes can include radiated gluons or quarks that recoil against the boson, which introduce non-zero boson transverse momentum, as illustrated in Figure 1.2. In $\mathrm{W} \to \ell\nu$ events, the particles visible to the detector are the lepton with momentum p^{ℓ} and the hadronic recoil against the W boson with momentum u. Similar to the $Z(\ell\ell)$+jets and γ+jets cases introduced in the last chapter, the transverse component $\vec{u}_{\mathrm{T}}$ is expected to balance the transverse momentum of the W boson ($\vec{p_{\mathrm{T}}}^{\mathrm{W}}$), i.e.,

$$\vec{p_{\mathrm{T}}}^{\mathrm{W}} = \vec{p_{\mathrm{T}}}^{\ell} + \vec{p_{\mathrm{T}}}^{\nu} = -\vec{u}_{\mathrm{T}}\,. \tag{8.1}$$

Equation 8.1 can be used to calculate the unknown $\vec{p_{\mathrm{T}}}^{\nu}$:

$$\vec{p_{\mathrm{T}}}^{\nu} = \vec{p}_{\mathrm{T}}^{\mathrm{miss}} = -(\vec{u}_{\mathrm{T}} + \vec{p_{\mathrm{T}}}^{\ell})\,. \tag{8.2}$$

The transverse mass m_{T} [123], defined as:

$$\frac{m_{\mathrm{T}}^2}{2} = p_{\mathrm{T}}^{\ell} p_{\mathrm{T}}^{\nu}(1 - \cos\delta\phi) = p_{\mathrm{T}}^{\ell} p_{\mathrm{T}}^{\nu}(1 - \frac{\vec{p_{\mathrm{T}}}^{\ell} \cdot \vec{p_{\mathrm{T}}}^{\nu}}{p_{\mathrm{T}}^{\ell} p_{\mathrm{T}}^{\nu}})\,, \tag{8.3}$$

will then become:

$$\frac{m_{\mathrm{T}}^2}{2} = p_{\mathrm{T}}^{\ell}(p_{\mathrm{T}}^{\ell} + |\vec{u}_{\mathrm{T}} + \vec{p_{\mathrm{T}}}^{\,\ell}|) + \vec{u}_{\mathrm{T}} \cdot \vec{p_{\mathrm{T}}}^{\,\ell}\,. \tag{8.4}$$

The transverse mass m_{T} can therefore be calculated from the measured $\vec{p_{\mathrm{T}}}^{\,\ell}$ and $\vec{u}_{\mathrm{T}}$.

The mass of the W boson, m_{W}, can be measured by comparing the p_{T}^{ℓ} and m_{T} distributions from experimental data with a set of templates obtained from the calibrated MC simulations produced assuming different m_{W} values: the template with the largest similarity should indicate the correct m_{W} within uncertainties. Figure 8.1 provides an example of the p_{T}^{ℓ} and m_{T} distribution comparisons assuming different m_{W} values.

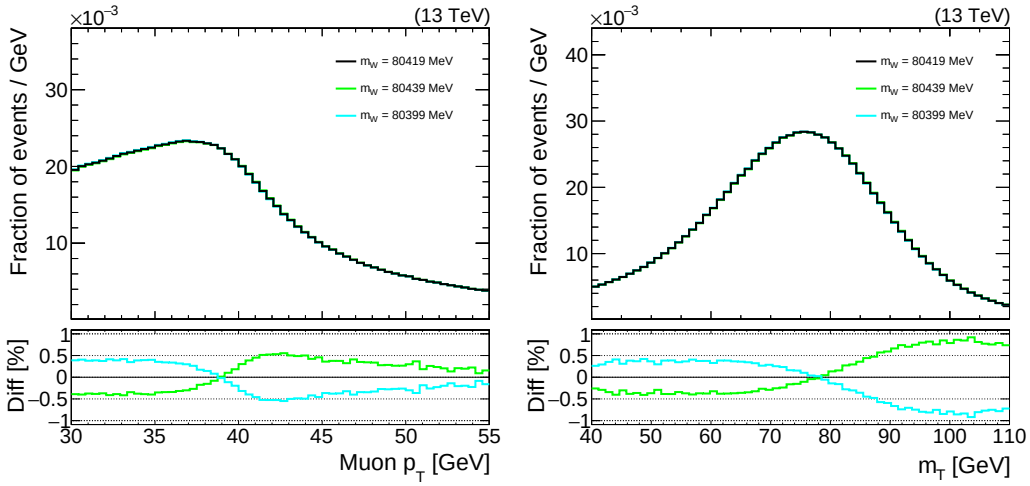


Figure 8.1: Distributions of the muon p_{T} (left) and the transverse mass m_{T} (right) with different m_{W} values. The relative difference in the bottom panel is calculates as the difference between the distribution assuming $m_{\mathrm{W}} = 80439(80399)\,\mathrm{MeV}$ and the distribution assuming $m_{\mathrm{W}} = 80419\,\mathrm{MeV}$, divided by the distribution assuming $m_{\mathrm{W}} = 80419\,\mathrm{MeV}$.

8.1.2 Transverse Mass m_{T} and Charged Lepton p_{T}

The mass of a charged lepton is negligible compared with the mass of the W boson. Therefore, in the W boson rest frame of a $\mathrm{W} \to \ell\nu$ decay, the charged lepton is emitted back-to-back with respect to the neutrino, and carry half of the W mass as its momentum: $p = m_{\mathrm{W}}/2$. The transverse momentum of the charged lepton is given by:

$$p_{\mathrm{T}} = p \sin\theta = \frac{1}{2} m_{\mathrm{W}} \sin\theta \,, \tag{8.5}$$

where θ is the polar angle of the charged lepton with respect to the beam pipe. The differential cross section in p_{T}, $d\sigma/dp_{\mathrm{T}}$, can be written as:

$$\begin{aligned} \frac{d\sigma}{dp_{\mathrm{T}}} &= \frac{d\sigma}{d\cos\theta} \frac{d\cos\theta}{dp_{\mathrm{T}}} \\ &= \frac{d\sigma}{d\cos\theta} \frac{2p_{\mathrm{T}}}{m_{\mathrm{W}}} \frac{d\cos\theta}{d\sin\theta} \\ &= \frac{d\sigma}{d\cos\theta} \frac{2p_{\mathrm{T}}}{m_{\mathrm{W}}} \frac{1}{\sqrt{m_{\mathrm{W}}^2/4 - p_{\mathrm{T}}^2}} \,, \end{aligned} \tag{8.6}$$

where the differential cross section $d\sigma/d\cos\theta$ follows

$$\frac{d\sigma}{d\cos\theta} = \sigma_0 (1 + \cos^2\theta) \tag{8.7}$$

Therefore, $d\sigma/dp_{\mathrm{T}}$ can be written as:

$$\frac{d\sigma}{dp_{\mathrm{T}}} = \sigma_0 \frac{m_{\mathrm{W}}^2 - 2p_{\mathrm{T}}^2}{m_{\mathrm{W}}^2} \frac{4p_{\mathrm{T}}}{m_{\mathrm{W}}} \frac{1}{\sqrt{m_{\mathrm{W}}^2/4 - p_{\mathrm{T}}^2}} \,, \tag{8.8}$$

which has a singularity at $p_{\mathrm{T}} = m_{\mathrm{W}}/2$. This divergence does not lead to an infinite cross section because the W boson mass follows a Breit-Wigner shape. However, there is still a strong peak with a sharply falling edge in the distribution, which occurs at half of m_{W}, referred to as the Jacobian edge, since it comes from the $d\cos\theta/d\sin\theta$ term. Similar structures can be derived for the transverse mass m_{T}, which peaks at m_{W} followed by a Jacobian edge. The black distributions in Figure 8.2 provide an illustration: the Jacobian edges exist in both m_{T} and charged lepton p_{T} distributions, with the first after the m_{W} value, and the second after the $m_{\mathrm{W}}/2$ value.

The m_{W} measurements from m_{T} and from p_{T}^{ℓ} provide a powerful cross-check because of their complementary uncertainties. The measurement from m_{T} is dominated by detector effects that affect the recoil measurement, such as pileup, reconstruction efficiency, and experimental resolutions. However, the measurement from m_{T} is more robust to the modeling of p_{T}^{W}. The measurement from p_{T}^{ℓ}, on the other hand, is mostly affected by the unknown p_{T}^{W} and the lepton momentum scale. As shown in Figure 8.2, the m_{T} distribution does not change significantly with finite p_{T}^{W} smearing, but changes dramatically with detector smearings. The opposite behavior is observed for p_{T}^{ℓ}. Combining the m_{W} measurement from the two distributions should bring the uncertainties on the m_{W} measurement to a minimum.

The recoil of the W boson is studied in the following sections, in order to find a recoil estimator that could help bring down the m_{W} uncertainties from the m_{T} fit.

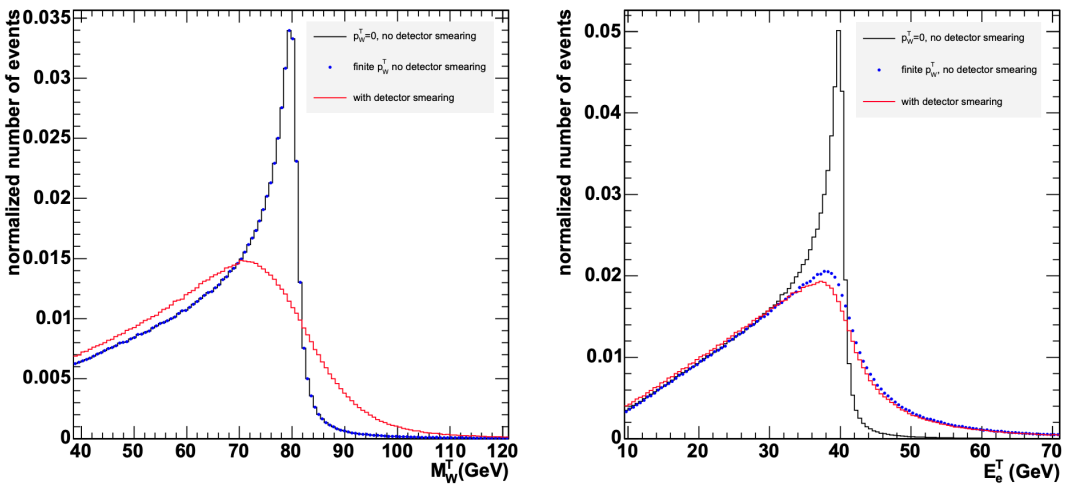


Figure 8.2: Distributions of the transverse mass (left) and the lepton transverse momentum (right) in $W \to \ell\nu$ decays. The three histograms in each plot are the distribution at generator level with $p_{\mathrm{T}}^{W} = 0$ (solid black), with finite p_{T}^{W} (dotted blue), and with experimental resolutions from the detector (solid red), respectively [124].

8.2 Generator-level Recoil Studies

The recoil analysis starts with generator-level studies. The simulated events are required to pass the single-muon selections in Section 7.2.2. Figure 8.3 shows the generator-level p_{T} and η distributions of the W boson after the single-muon selections. (In the later sections $p_{\mathrm{T}}^{\mathrm{W}}$ may be represented with q_{T}, to be consistent with the previous DeepMET chapter.) The q_{T} distribution peaks around 5 GeV, and falls exponentially as q_{T} goes higher, while the η distribution peaks around ± 2.5.

8.2.1 From Generator-level Recoil to Transverse Mass

Figure 8.4 shows the m_{T} distributions calculated using the reconstructed p_{T}^{μ} and the generator-level q_{T}, in different q_{T} bins. The distributions follow a similar pattern in each q_{T} bin: they peak at the m_{W} value with a sharply falling Jacobian

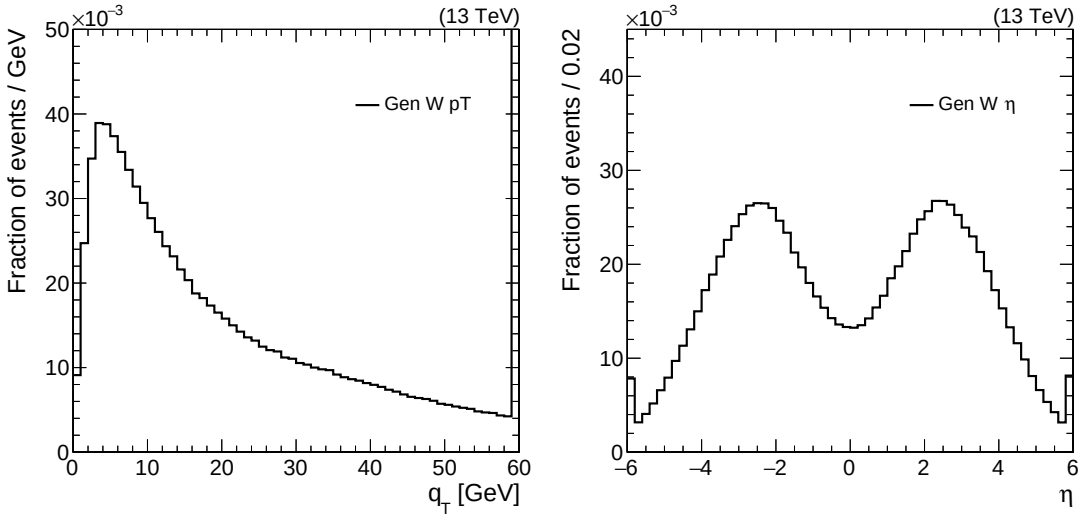


Figure 8.3: Distributions of the generator-level p_{T} (left) and η (right) of the W boson after the single-muon selections. The underflow (overflow) contents are included in the first (last) bin.

edge.

8.2.2 Generator-level Final-state Particles

The fact that q_{T} peaks at 5 GeV and then falls exponentially implies that in many events the hadronic recoil against the W boson consists of rather soft activities not clustered into high-p_{T} jets. Therefore, the proper evaluation of the hadronic recoil needs to focus on the recoiled individual particles, either stable particles at the generator level, or PF candidates at the reconstruction level.

Figure 8.5 shows the p_{T} (top), rapidity y (bottom left) and pseudorapidity η (bottom right) distributions of all generator-level stable particles in the events. All distributions are normalized to the total number of events. Most of the particles are soft and centrally distributed. Table 8.1 shows the average number of different particles per event.

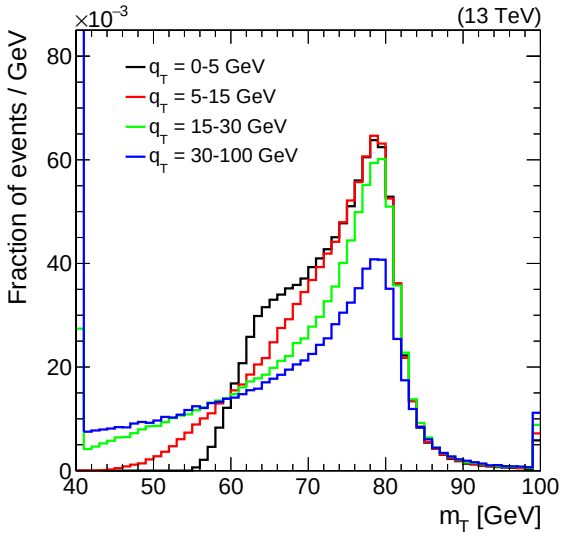


Figure 8.4: Distributions of m_{T} in different q_{T} ranges, calculated using the reconstructed p_{T}^{μ} and the generator-level q_{T}, after the single muon selections. The underflow (overflow) contents are included in the first (last) bin. All of the distributions are normalized to unity.

Table 8.1: Average number of generator-level stable particles per $W \to \mu\nu$ event, after the single-muon selections.

	All	$p_{\mathrm{T}} > 0.5\,\mathrm{GeV}$	Fraction
Charged particles	149.2	69.0	0.62
Photons	147.5	26.0	0.23
Neutral hadrons	25.8	16.7	0.15

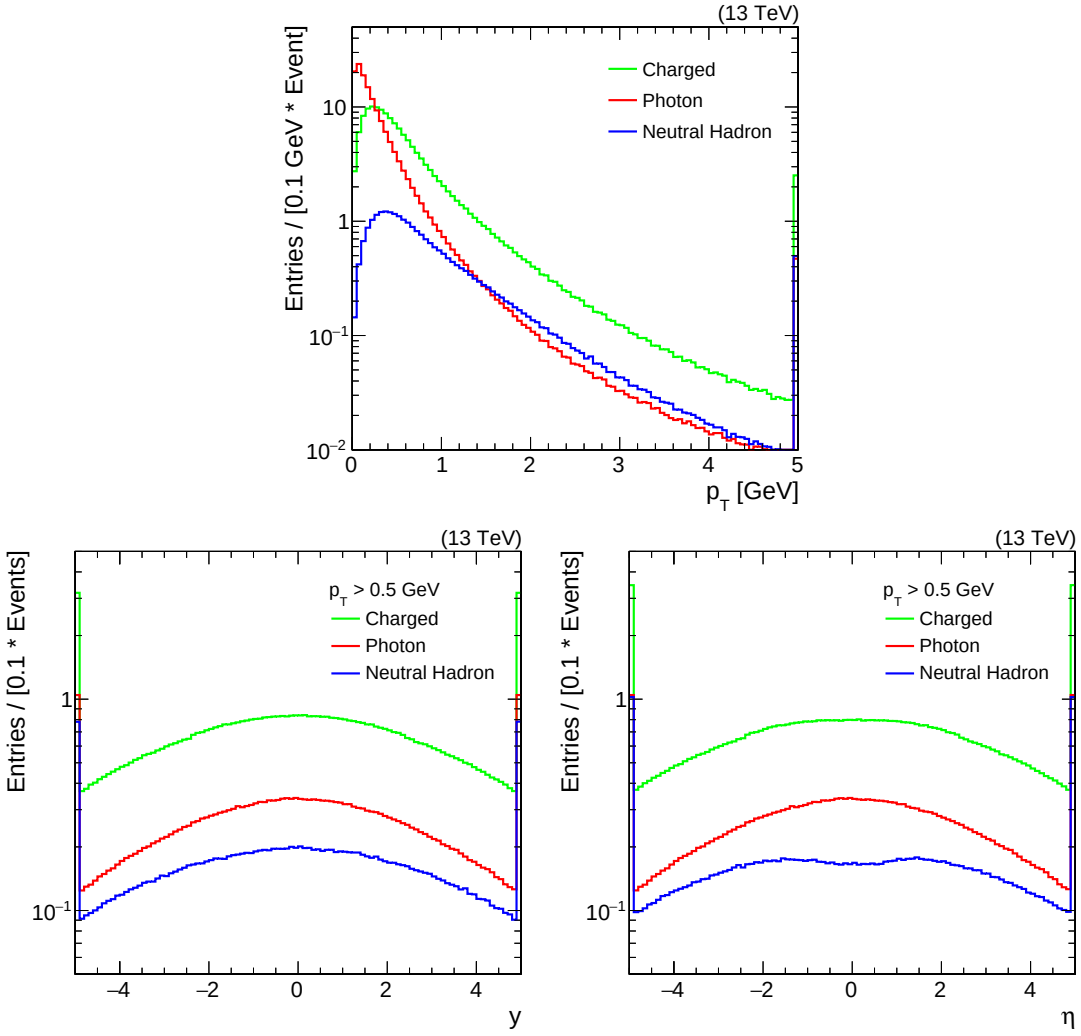


Figure 8.5: Distributions of the p_{T} (top), rapidity y (bottom left), and η (bottom right) of all generator-level stable particles per event, after the single-muon selections. In the bottom plots, particles are required to pass the $p_{\mathrm{T}} > 0.5$ GeV selection. The underflow (overflow) contents are included in the first (last) bin.

8.2.3 Generator-level Recoil Estimators

Figure 8.6 shows the distributions of the recoil responses after applying different selections on the generator-level recoiled stable particles. As shown in the plot, the response (solid black) reaches a value close to unity at high q_{T} within the fiducial coverage of the CMS detector ($|\eta| < 5.0$). At low q_{T}, instead, where most of W events reside, the response is around 90%. If only the recoiled particles with p_{T} above 0.5 GeV are selected, the response drops to around 80% at low q_{T}. If the HF region is ignored, i.e., in the $|\eta| < 3.0$ region, the response (blue) drops to 60% - 90%. For the recoil using only charged particles within the tracker volume ($|\eta| < 2.5$), the response is around 35% in the low-q_{T} region, and goes up to 50% in the high-q_{T} region; the $p_{\mathrm{T}} > 0.5$ GeV cut causes the estimator based on charged particles to drop slightly in the low-q_{T} region.

Figure 8.7 shows the resolutions of $u_{\parallel}$ (left) and $u_{\perp}$ (right) as a function of q_{T}. Figure 8.8 shows the same distributions but with the response-corrected resolutions. Table 8.2 presents the overall resolutions of $u_{\parallel}$ and $u_{\perp}$ (with and without response corrections) with different selections. In the optimal case, where all recoiled particles within the CMS acceptance ($|\eta| < 5.0$) have been successfully reconstructed, perfectly measured, and without any contamination from other sources, the resolutions of $u_{\parallel}$ and $u_{\perp}$ are around 4 GeV, across the whole q_{T} range. Without the information from the forward region, the response-corrected resolution of $u_{\perp}$ is about 10 GeV. While using only the tracking information, the response-corrected resolution of $u_{\perp}$ increases to 15 GeV.

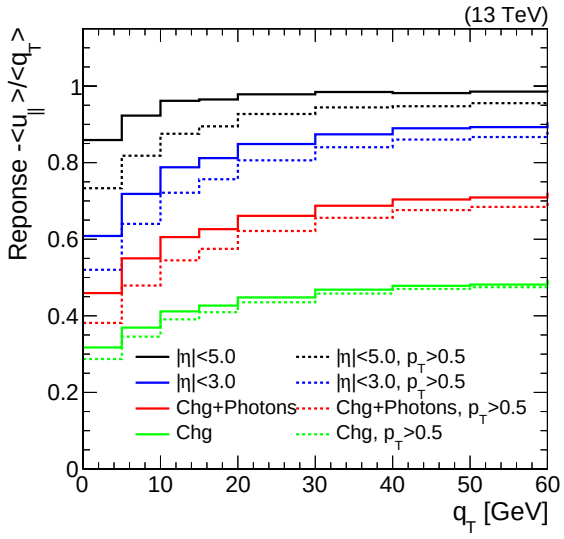


Figure 8.6: Distributions of recoil responses using generator-level stable particles with different selections. The solid black is the recoil using all these particles with $|\eta| < 5.0$ (the maximum coverage of the CMS detector). The solid blue is using all these particles with $|\eta| < 3.0$ (excluding CMS HF and forward subdetectors). The solid red is using all charged particles within $|\eta| < 2.5$ and photons within $|\eta| < 3.0$ (CMS Tracker + ECAL). The solid green is using all stable charged particles within $|\eta| < 2.5$ (CMS Tracker). The dashed distributions are applying the corresponding solid selections plus the p_{T} >0.5 GeV cut.

Table 8.2: Resolutions of $u_{\parallel}$ and $u_{\perp}$ in W($\mu\nu$) MC events with and without response corrections using generator-level stable particles with different selections, after the single-muon selections.

Selections on generator-level particles (p_{T} unit: GeV)	$\sigma(u_{\parallel})$ [GeV]	Corrected $\sigma(u_{\parallel})$ [GeV]	$\sigma(u_{\perp})$ [GeV]	Corrected $\sigma(u_{\perp})$ [GeV]
$\lvert\eta\rvert < 5.0$	4.1	4.3	4.0	4.2
$\lvert\eta\rvert < 5.0, p_{\mathrm{T}} > 0.5$	4.6	5.3	4.5	5.1
$\lvert\eta\rvert < 3.0$	8.3	10.7	7.7	9.7
$\lvert\eta\rvert < 3.0, p_{\mathrm{T}} > 0.5$	8.1	11.1	7.3	9.9
charged $\lvert\eta\rvert < 2.5$, photon $\lvert\eta\rvert < 3.0$	11.1	16.4	7.9	12.8
charged $\lvert\eta\rvert < 2.5$, photon $\lvert\eta\rvert < 3.0$, both with $p_{\mathrm{T}} > 0.5$	11.0	17.1	7.4	13.1
charged $\lvert\eta\rvert < 2.5$	14.2	20.8	6.6	15.6
charged $\lvert\eta\rvert < 2.5$, $p_{\mathrm{T}} > 0.5$	14.3	21.2	6.4	15.8

The recoil using only charged particles has a small $u_{\perp}$ resolution of around 6 GeV without the response corrections. However, due to its low response, the $u_{\perp}$ resolution becomes above 15 GeV after the response correction.

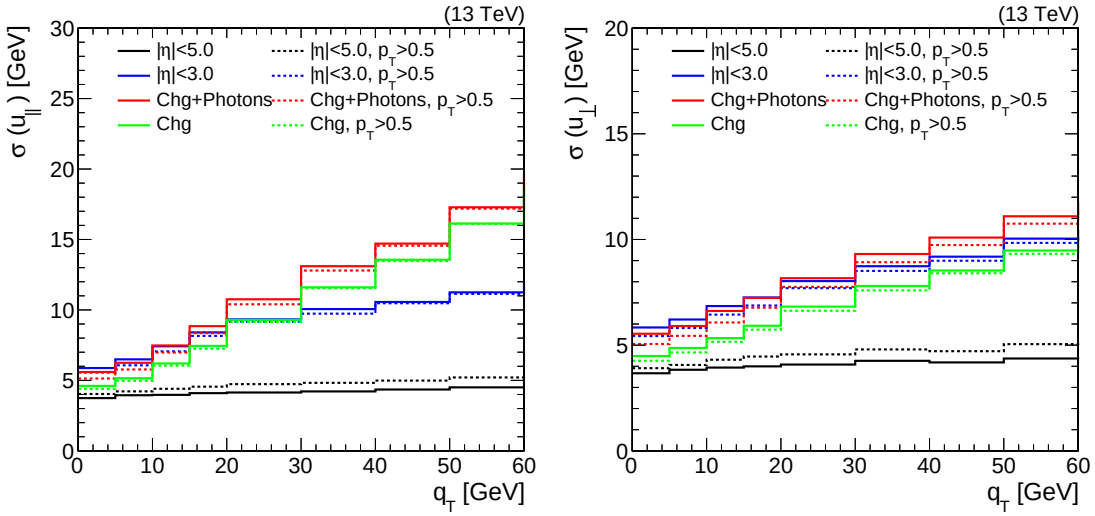


Figure 8.7: Resolutions of $u_{\parallel}$ (left) and $u_{\perp}$ (right) *vs.* q_{T} using generator-level final-state particles with different selections in W($\mu\nu$) MC, after the single muon selections.

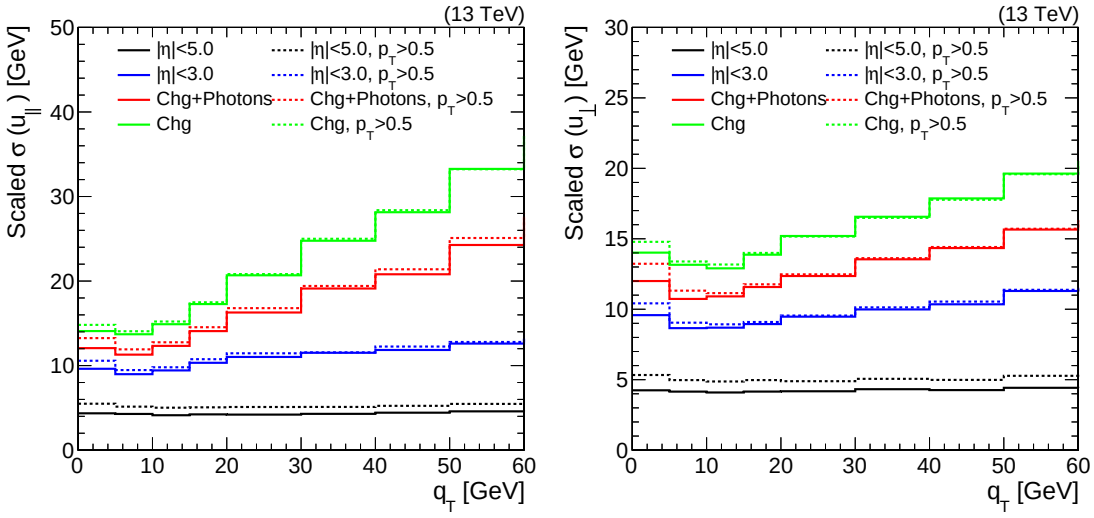


Figure 8.8: Resolutions of $u_{\parallel}$ (left) and $u_{\perp}$ (right) *vs.* q_{T} using generator-level final-state particles with different selections in W($\mu\nu$) MC, after the single muon selections. The resolutions are corrected with the responses taken from Figure 8.6.

8.3 Reconstruction-level Recoil Studies

The recoil estimators calculated from reconstructed PF candidates are discussed in this section. Compared with the results using generator-level stable particles, the responses and resolutions of recoil estimators using reconstructed PF candidates are affected by the pileup interactions, reconstruction efficiencies and the resolution of momentum measurements.

8.3.1 Reconstructed PF Candidates

Figure 8.9 shows the averaged p_{T} and η distributions of all reconstructed PF candidates per event, after the single-muon selections. Most of the particles are soft and centrally distributed. Table 8.3 includes the average number of different particles per event. Compared with the generator-level stable particles shown in

Table 8.1, the numbers have increased 3-5 times, mainly because the generator-level information does not include pileup particles.

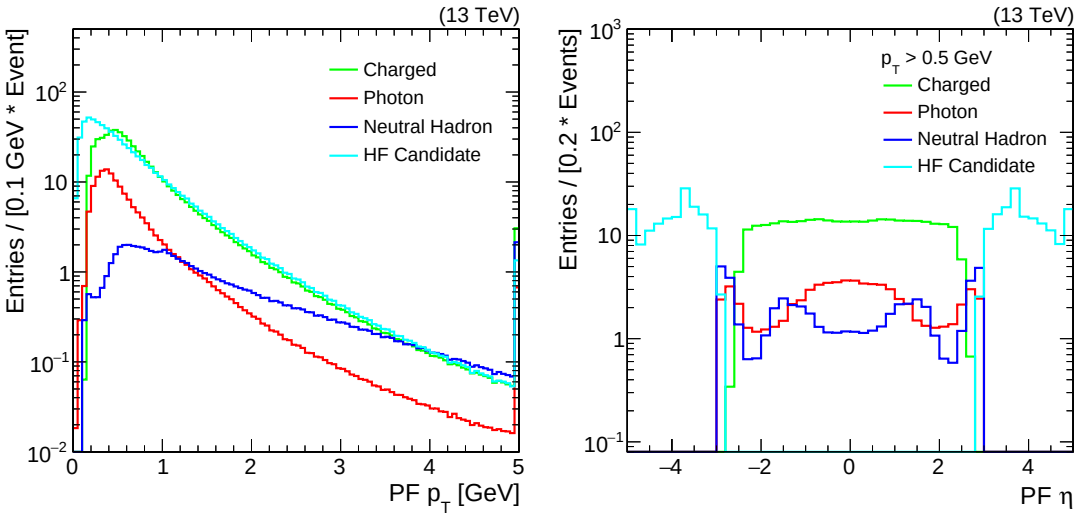


Figure 8.9: Distribution of the p_{T} and η of all reconstructed PF candidates per $\mathrm{W} \to \mu\nu$ MC event, after the single-muon selections. In the right plot, particles are required to pass the $p_{\mathrm{T}} > 0.5\,\mathrm{GeV}$ selection. The underflow (overflow) contents are included in the first (last) bin.

Table 8.3: Average number of reconstructed PF candidates per $\mathrm{W} \to \mu\nu$ MC event, after the single-muon selections.

	All	$p_{\mathrm{T}} > 0.5\,\mathrm{GeV}$	Fraction
Charged	540.0	334.1	0.43
Photon	147.5	71.4	0.09
Neutral hadrons	61.3	55.0	0.07
HF candidates	699.7	315.7	0.41

8.3.2 Recoil Estimators

Similar to Table 8.2, Table 8.4 presents the resolution values of $u_{\parallel}$ and $u_{\perp}$ (with and without response corrections) with different selections of PF Candidates. The PUPPI weights have been included in all of the recoil calculations, because the same selections without PUPPI weights always lead to worse performances due to

Table 8.4: Resolutions of $u_\parallel$ and $u_\perp$ in W($\mu\nu$) MC events with and without response corrections using PF candidates with different selections, after the single muon selections. The PUPPI weights are included in the recoil calculations.

Selections on PF candidates (p_T unit in GeV)	$\sigma(u_\parallel)$ [GeV]	Corrected $\sigma(u_\parallel)$ [GeV]	$\sigma(u_\perp)$ [GeV]	Corrected $\sigma(u_\perp)$ [GeV]
$\lvert\eta\rvert < 5.0$	13.1	20.2	11.3	18.3
$\lvert\eta\rvert < 5.0, p_\mathrm{T} > 0.5$	13.0	20.5	11.1	18.5
$\lvert\eta\rvert < 3.0$	12.7	20.1	10.2	17.3
$\lvert\eta\rvert < 3.0, p_\mathrm{T} > 0.5$	12.6	20.4	10.1	17.4
charged $\lvert\eta\rvert < 2.5$	15.2	23.7	6.7	18.0
charged $\lvert\eta\rvert < 2.5$, $p_\mathrm{T} > 0.5$	15.2	24.0	6.6	18.3

pileup effects.

Differently from what is observed with generator-level stable particles reported in Table 8.2, in the case of PF candidates the response-corrected resolutions of $u_\perp$ are similar with different selections, while for $u_\parallel$, including neutral particles within $|\eta| < 3.0$ helps the recoil measurement, extending it to the HF region ($3.0 < |\eta| < 5.0$) does not improve the resolution. This indicates that the benefit from including reconstructed forward particles to the W recoil measurement is limited, probably due to pileup effects and the worse resolutions of the momentum magnitude and direction. It shall also be noted that the p_T cut at 0.5 GeV has little impact on the resolutions.

Table 8.5 lists the resolutions of PF $p_\mathrm{T}^\mathrm{miss}$, PUPPI $p_\mathrm{T}^\mathrm{miss}$ and DeepMET in the same W($\mu\nu$) MC events, where the PF $p_\mathrm{T}^\mathrm{miss}$ and PUPPI $p_\mathrm{T}^\mathrm{miss}$ are after the Type-I corrections. The results are similar to the Z($\mu\mu$) results shown in Table 7.2, as expected. DeepMET maintains the smallest resolutions among all the estimators.

Figure 8.10 shows the comparisons of the q_T (left), $p_\mathrm{T}^\mathrm{miss}$ (center) and the transverse mass m_T (right) reconstructed using PF $p_\mathrm{T}^\mathrm{miss}$, PUPPI $p_\mathrm{T}^\mathrm{miss}$ and DeepMET,

Table 8.5: Resolutions of $u_{\parallel}$ and $u_{\perp}$ of PF $p_{\mathrm{T}}^{\mathrm{miss}}$, PUPPI $p_{\mathrm{T}}^{\mathrm{miss}}$ and DeepMET with and without response corrections in W($\mu\nu$) MC, after the single-muon selections.

Type	$\sigma(u_{\parallel})$ [GeV]	Corrected $\sigma(u_{\parallel})$ [GeV]	$\sigma(u_{\perp})$ [GeV]	Corrected $\sigma(u_{\perp})$ [GeV]
PF $p_{\mathrm{T}}^{\mathrm{miss}}$	21.3	25.0	20.8	24.4
PUPPI $p_{\mathrm{T}}^{\mathrm{miss}}$	15.7	21.7	14.3	20.0
DeepMET	13.5	17.1	12.1	15.4

in which DeepMET has the sharpest peak in $p_{\mathrm{T}}^{\mathrm{miss}}$ and m_{T}. Also note that the response of PUPPI $p_{\mathrm{T}}^{\mathrm{miss}}$ is smaller than DeepMET and PF $p_{\mathrm{T}}^{\mathrm{miss}}$ in the low-q_{T} region, as shown in Figure 7.8, which would make the $p_{\mathrm{T}}^{\mathrm{miss}}$ and m_{T} distribution artificially better. The most powerful metric, after all, should be the systematic uncertainty on m_{W} evaluated by varying the W p_{T} distribution measured from different recoil estimators, which will be discussed in the later sections.

8.4 Data-MC Comparisons of DeepMET in the W Signal Region

The data-MC agreement of DeepMET after the single-muon selections is evaluated in this section.

8.4.1 Background Estimations

Most of the events in data are from $\mathrm{W} \to \mu\nu$ process after the single muon selections, with backgrounds from two major topologies:

- isolated lepton backgrounds: they arise from electroweak processes, for example, the Drell-Yan ($\mathrm{Z}/\gamma^{*} \to \ell\ell$) process; or the top quark pair production, when the top quarks decay to $\mathrm{W} + \mathrm{b}$ is followed by a leptonic decay of the W

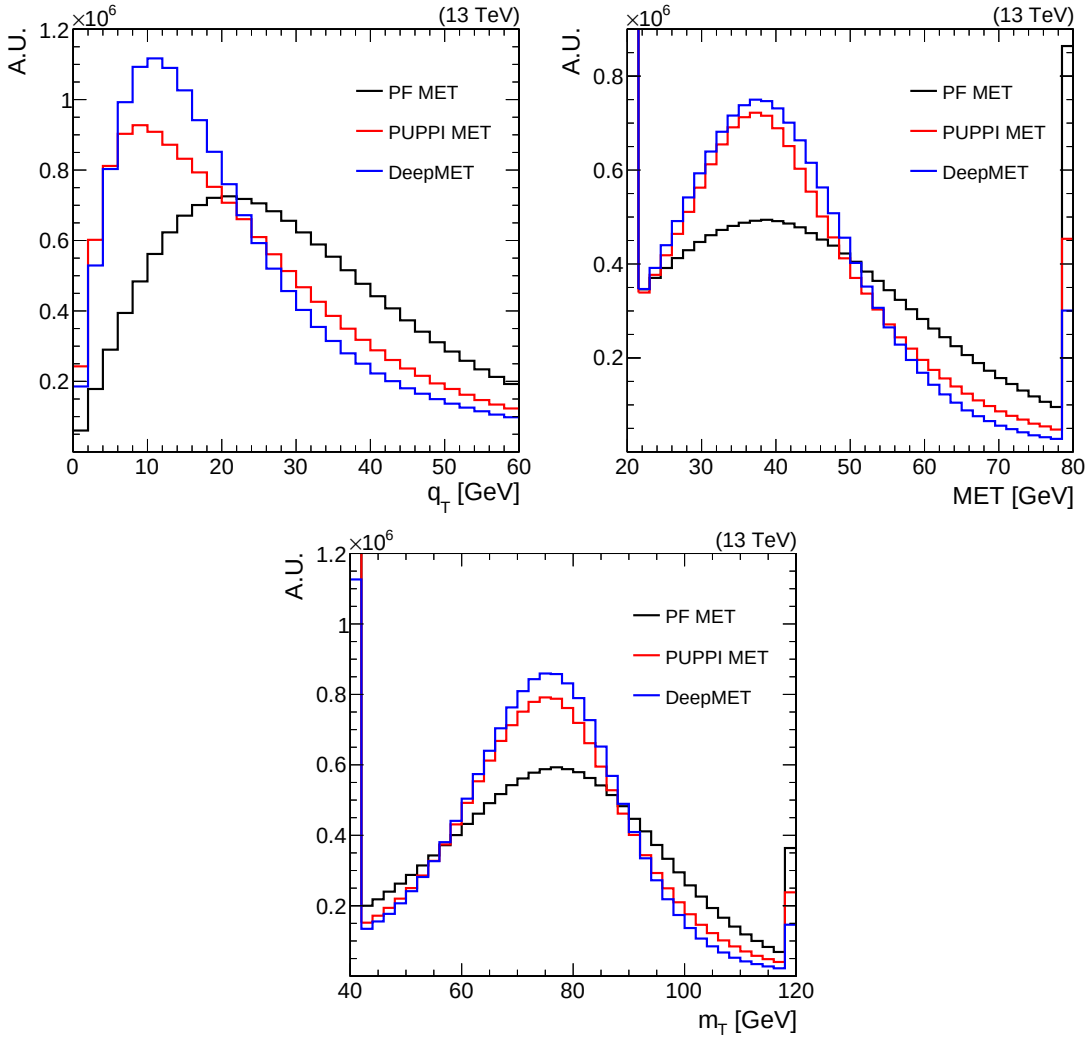


Figure 8.10: Comparisons of q_{T} (top left), $p_{\mathrm{T}}^{\mathrm{miss}}$ (top right) and transverse mass m_{T} (bottom) distributions in W($\mu\nu$) MC, reconstructed using PF MET (black), PUPPI MET (red) and DeepMET (blue), after the single-muon selections. The underflow (overflow) contents are included in the first (last) bin.

boson; contributions from the dibosons (WW, WZ, ZZ) are small because of the small production cross sections of these processes.

- non-isolated lepton backgrounds: they are dominated by the QCD multijet production, where the non-isolated lepton comes principally from the semileptonic decay of an heavy-flavor jets.

The contributions from misidentified muons are small considering the excellent muon identifications.

For the isolated lepton backgrounds, the kinematics of these electroweak processes are well simulated in MC and their cross sections are well measured. Therefore, these backgrounds are estimated directly using the MC simulations.

The non-isolated QCD background is estimated with a data-driven method. In the muon object selections, the muon candidate is required to have the PF isolation smaller than 0.15 (signal region). We invert this requirement and require one muon with PF isolation between 0.3 and 0.6, in order to create a control region dominated by QCD events. Distributions of q_{T}, $p_{\mathrm{T}}^{\mathrm{miss}}$ and the m_{T} calculated from DeepMET in this region are shown in Figure 8.11, in which the QCD contributions are taken from MC simulations and plotted only for illustration purposes. For the data-driven QCD estimation, the contributions of the electroweak processes containing a genuine muon from a W and Z decay are subtracted from data in the control region, using the MC simulations. The obtained shape is used as an estimator of the QCD background shape in the signal region. The normalization of this background is obtained by comparing the observed and expected number of events in the region

with $m_{\mathrm{T}} < 30\,\mathrm{GeV}$ and PF Isolation smaller than 0.15.

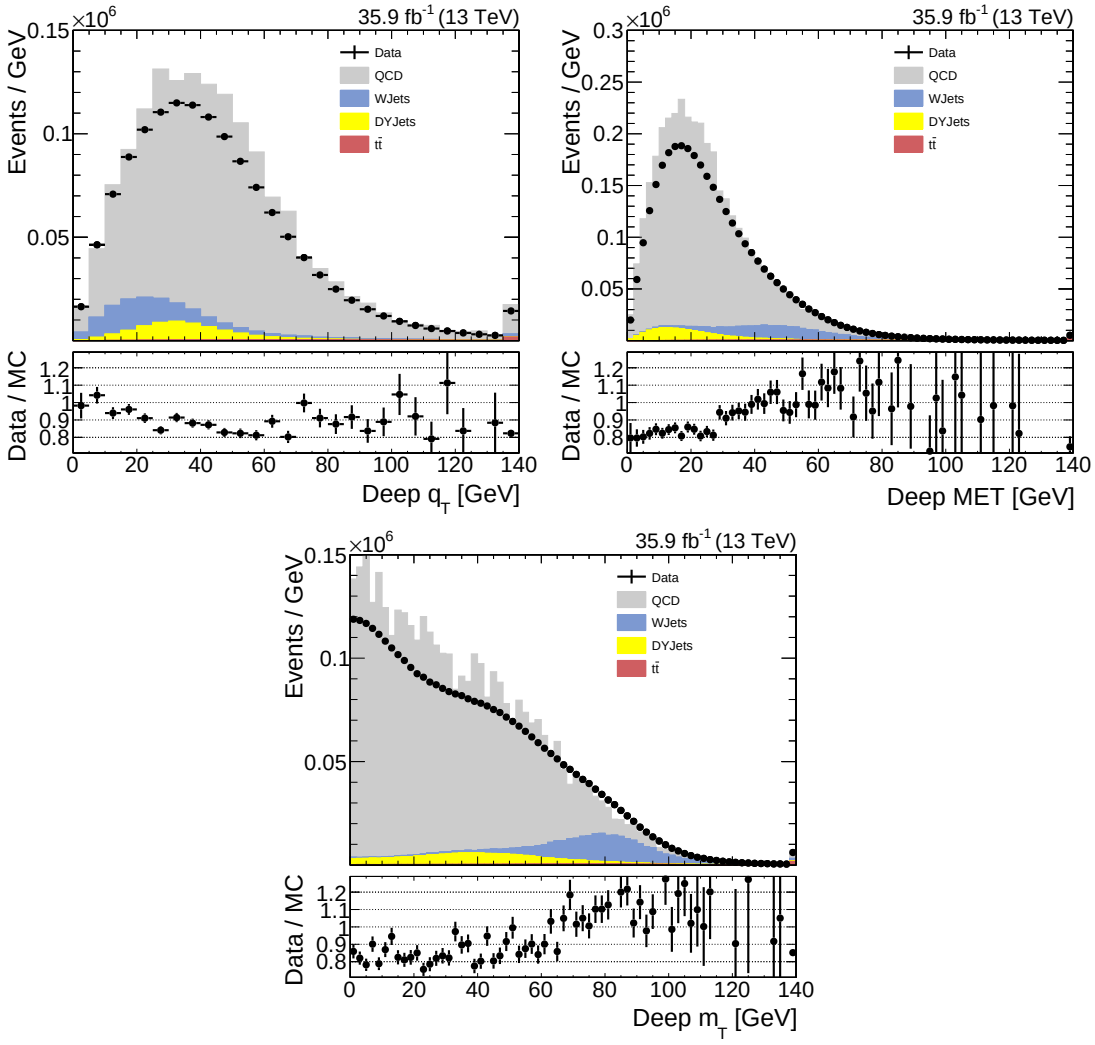


Figure 8.11: Distributions of q_{T} (top left), $p_{\mathrm{T}}^{\mathrm{miss}}$ (top right) and m_{T} (bottom) calculated from DeepMET in the QCD dominated control region. The QCD distribution is taken from MC simulations, for the illustration purpose only. The underflow (overflow) contents are included in the first (last) bin.

The cut values on the muon PF isolations for the control region, 0.3 and 0.6, are verified by using different cut values (0.3 and 0.45, 0.45 and 0.6), and the impacts on the final QCD background shape are found to be small. The normalization of the QCD background can be also treated as a free-floating parameter, and estimated from fits of different distributions. In this dissertation, it is normalized to the region

with $m_{\mathrm{T}} < 30\,\mathrm{GeV}$, with the purpose of carrying out a fast data-MC comparison. It is also verified by varying the m_{T} cut values, with only small differences found.

8.4.2 Data-MC Comparisons of DeepMET

With the background estimation methods described above, Figure 8.12 shows the data-MC comparisons of q_{T} (top left), $p_{\mathrm{T}}^{\mathrm{miss}}$ (top right) and m_{T} (bottom) calculated from DeepMET in the events after the single muon selections. Figure 8.13 shows the same set of plots, but after the recoil corrections described in Section 7.6.1.

The comparisons of Figures 8.12 and 8.13 demonstrate that the corrections significantly improve the data-MC agreement at high m_{T}. The agreement is within the typical W+jets theoretical uncertainties (about 5%), well suited for any general-purpose analysis.

In a future m_{W} analysis, the backgrounds, especially the non-isolated muon backgrounds from QCD, can be further reduced with $p_{\mathrm{T}}^{\mathrm{miss}}$ and m_{T} cuts. No $p_{\mathrm{T}}^{\mathrm{miss}}$ or m_{T} cut is applied here to demonstrate the overall good data-MC agreement. The background estimations can also be improved with more precise methods, for example, measuring the misidentification rate of non-isolated muons identified as isolated in bins of muon p_{T} and η. These could fix the minor shape differences observed in the QCD dominated region in these figures. The muon momentum scale can be calibrated and measured more precisely to improve the agreement.

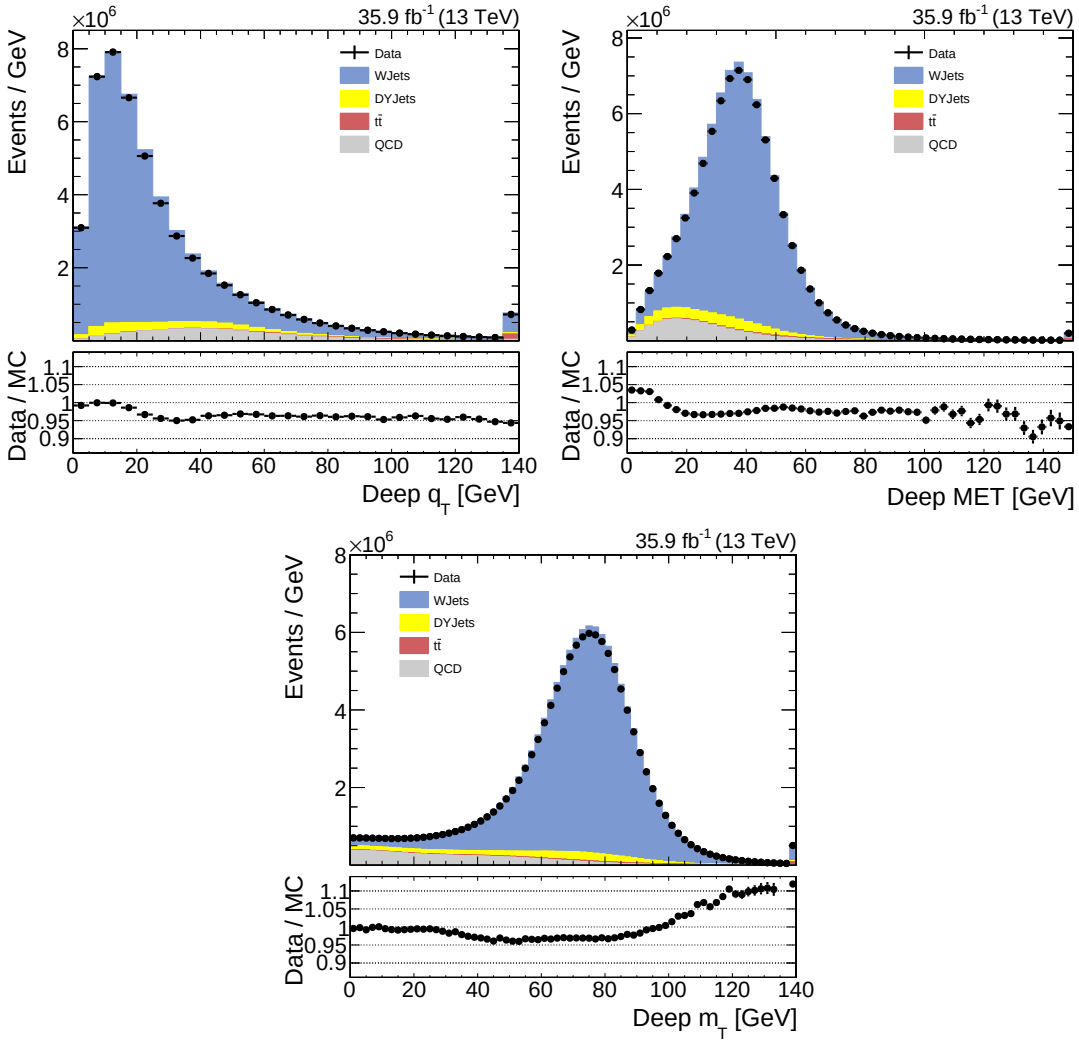


Figure 8.12: Data-MC comparisons of q_{T} (top left), $p_{\mathrm{T}}^{\mathrm{miss}}$ (top right), and m_{T} (bottom) calculated from DeepMET in events after the single-muon selections. The underflow (overflow) contents are included in the first (last) bin.

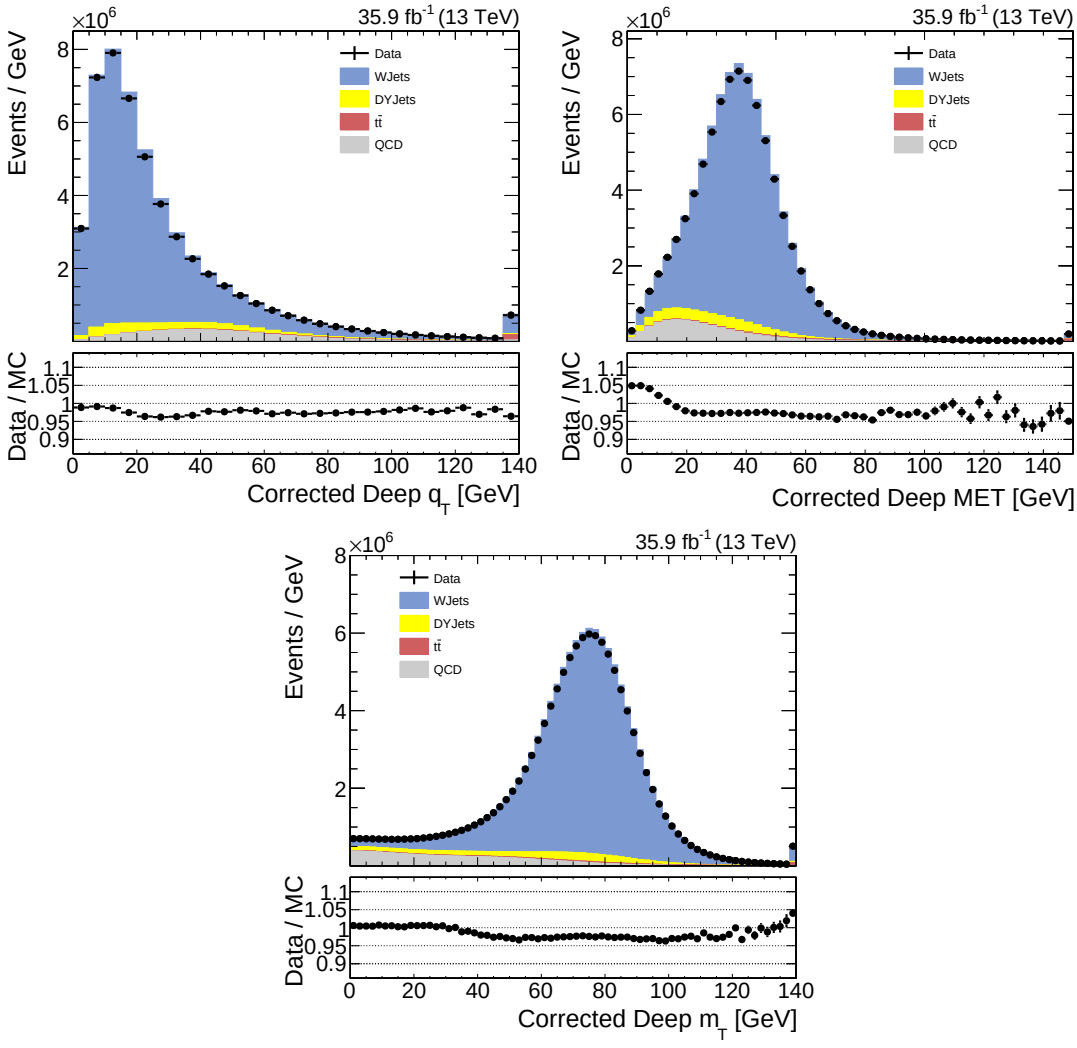


Figure 8.13: Data-MC comparisons of q_{T} (top left), $p_{\mathrm{T}}^{\mathrm{miss}}$ (top right), and m_{T} (bottom) calculated from DeepMET in events after the single-muon selections. The recoil corrections described in Section 7.6.1 are applied. The underflow (overflow) contents are included in the first (last) bin.

8.5 Uncertainties on W Mass Fit from Recoil

As discussed earlier, the W boson mass m_{W} is Lorentz invariant, and does not change with $p_{\mathrm{T}}^{\mathrm{W}}$. On the other hand, p_{T}^{ℓ} depends on the $p_{\mathrm{T}}^{\mathrm{W}}$ distribution. Therefore using p_{T}^{ℓ} to measure m_{W} requires a good understanding and modeling of $p_{\mathrm{T}}^{\mathrm{W}}$. The idea of using the recoil u_{T} and building m_{T} is to partially recover the lack of information on $p_{\mathrm{T}}^{\mathrm{W}}$ and make the m_{W} measurement less dependent on it. This is also shown in Figure 8.2.

In this section, the systematic uncertainty on m_{W} from the $p_{\mathrm{T}}^{\mathrm{W}}$ modeling are evaluated, in order to quantify the performance of different recoil estimators. The m_{W} fitting procedure is firstly introduced, then variations of the $p_{\mathrm{T}}^{\mathrm{W}}$ shape are briefly discussed, and finally the shifts of m_{W} from the $p_{\mathrm{T}}^{\mathrm{W}}$ shape variations are presented.

8.5.1 W Mass Fitting Procedure

The m_{W} value is extracted with a template fitting procedure as mentioned earlier. It consists in the comparison of the distributions of one variable (m_{T}, p_{T}^{ℓ} or $p_{\mathrm{T}}^{\mathrm{miss}}$, etc.) between the data and several MC templates built using different m_{W} hypotheses. For each hypothetical M_{new}, a new set of event weights will be generated using the following ratio of Breit-Wigner functions:

$$w = \frac{(M^2 - M_0^2)^2 + (\Gamma M_0)^2}{(M^2 - M_{new}^2)^2 + (\Gamma M_{new})^2}\,, \tag{8.9}$$

where M is the generated W mass in one event, M_0 and Γ are the default W mass and W width. This expression of event weights is an approximation because it does not take into account the relations between the mass and other parameters, such as Γ, the couplings, etc. However, it holds under the assumption $\Delta M = M_{new} - M_0 \ll \Gamma$, and fully satisfies the m_{W} analysis since the region of interest is $\Delta M < 100\,\mathrm{MeV}$.

The m_{T} distribution is used to extract the m_{W} in this section, in order to study different recoil estimators. The special case $u_{\mathrm{T}} = 0$ corresponds to $m_{\mathrm{T}} = 2p_{\mathrm{T}}^{\ell}$. The M_0 and Γ are taken from the MC@NLO settings used in the MC generation: $M_0 = 80.419\,\mathrm{GeV}$ and $\Gamma = 2.050\,\mathrm{GeV}$. The M_{new} parameters are set to be between $80.369\,\mathrm{GeV}$ and $80.469\,\mathrm{GeV}$; the event weights are generated for different M_{new} values every $5\,\mathrm{MeV}$. The 21 sets of m_{T} templates from the reweighted MC are morphed with `RooMomentMorph` provided by ROOT. The Asimov data [125] corresponds to the MC produced with $m_{\mathrm{W}} = M_0 = 80.419\,\mathrm{GeV}$, scaled to match the integrated luminosity of the data sample, $35.9\,\mathrm{fb}^{-1}$.

8.5.2 Event Selections and m_{T} Fit Range

In the ATLAS m_{W} analysis [122], events are required to pass the single-lepton selections, have $p_{\mathrm{T}}^{\mathrm{miss}}$ larger than $30\,\mathrm{GeV}$ and a measured recoil u_{T} smaller than $30\,\mathrm{GeV}$. The purposes of the last two selection criteria is to reduce the QCD multijet background contribution and minimize the model uncertainty on W boson productions at high transverse momentum. In that analysis, the upper boundary of the m_{T} fit range is varied between $[90,\ 100]\,\mathrm{GeV}$ and the lower boundary is varied

between [65, 70] GeV. The fit range is finally chosen to be $66 < m_{\mathrm{T}} < 99\,\mathrm{GeV}$, as this is found out to be the one with smallest total uncertainty in m_{W}.

To mimic the ATLAS event selections, in this section the events are required to pass the single muon selections listed in Section 7.2.2, and have $p_{\mathrm{T}}^{\mathrm{miss}} > 30\,\mathrm{GeV}$ and $u_{\mathrm{T}} < 30\,\mathrm{GeV}$, with the $p_{\mathrm{T}}^{\mathrm{miss}}$ and u_{T} calculated from the recoil estimator being studied. Figure 8.14 shows the reconstructed m_{T} distributions for different recoil estimators after the event selections. As discussed previously, PUPPI $p_{\mathrm{T}}^{\mathrm{miss}}$ and DeepMET have smaller resolutions compared with PF $p_{\mathrm{T}}^{\mathrm{miss}}$, and therefore their m_{T} distributions fall more quickly than the m_{T} reconstructed using PF $p_{\mathrm{T}}^{\mathrm{miss}}$. In the low u_{T} region, DeepMET has a larger response and a similar resolution compared with PUPPI $p_{\mathrm{T}}^{\mathrm{miss}}$. Therefore, the m_{T} difference between DeepMET and PUPPI $p_{\mathrm{T}}^{\mathrm{miss}}$ is small, but it will show up in Section 8.5.4 where the W p_{T} systematic uncertainty is studied.

In this section, the lower bound of m_{T} is chosen to be 65 GeV, the same value as the ATLAS lower bound; the upper bound of the m_{T} fit range is set to the 90% quantile of the m_{T} distribution, similarly to previous W recoil studies [126, 127]. This definition of the upper bound allows for a more sensible comparison of the different recoil estimators, each of which has a different resolution. It is found that the uncertainties on m_{W} from the m_{T} fit depends significantly on the m_{T} fit range; however, the ordering in performance of recoil estimators does not change when varying the m_{T} upper bound cut in 70%-95% quantiles. Table 8.6 lists the m_{T} fit range of different recoil estimators chosen with this method.

The above event selections and the fit range are a simplified setup, as our main

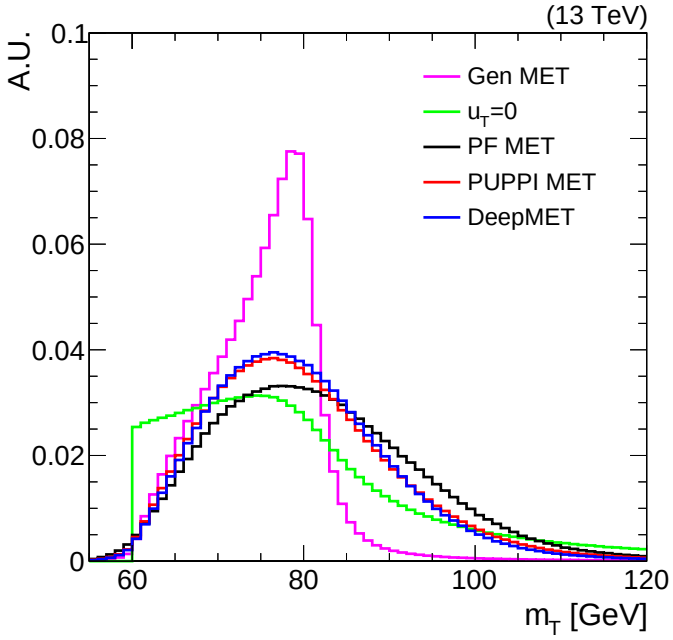


Figure 8.14: Distributions of reconstructed m_{T} for different recoil estimators in W+jets MC after the event selections. The distribution of $u_{\mathrm{T}} = 0$ is to set the u_{T} always to zero, in which case $m_{\mathrm{T}} = 2p_{\mathrm{T}}^{\ell}$.

Table 8.6: m_{T} fit ranges for different recoil estimators. The lower bound is fixed at 65 GeV, and the upper bound is set to the 90% quantile of the m_{T} distributions.

Recoil	min m_{T} [GeV]	max m_{T} (q_{90}) [GeV]
Generator-level $p_{\mathrm{T}}^{\mathrm{W}}$	65	83
PF $p_{\mathrm{T}}^{\mathrm{miss}}$	65	101
$u_{\mathrm{T}} = 0$	65	110
PUPPI $p_{\mathrm{T}}^{\mathrm{miss}}$	65	96
DeepMET	65	93

goal here is to vary the $p_{\mathrm{T}}^{\mathrm{W}}$ distributions and check how robust the m_{T} fit would be from different recoil estimators. In a future CMS m_{W} measurement, once the methods to estimate all systematic uncertainties are finalized, the event selections and fit ranges could be fully optimized to minimize the m_{W} uncertainties.

8.5.3 Uncertainties on the $p_{\mathrm{T}}^{\mathrm{W}}$ Modeling

Theoretical uncertainties on the $p_{\mathrm{T}}^{\mathrm{W,Z}}$ modeling are around a few percent. Figure 8.15 and 8.16 are from the latest ATLAS [128] and CMS [129] comparisons of $p_{\mathrm{T}}^{\ell\ell}$ distributions between different MC generators and data at $\sqrt{s} = 13\,\mathrm{TeV}$, where the ratios of MC over data vary within 10%.

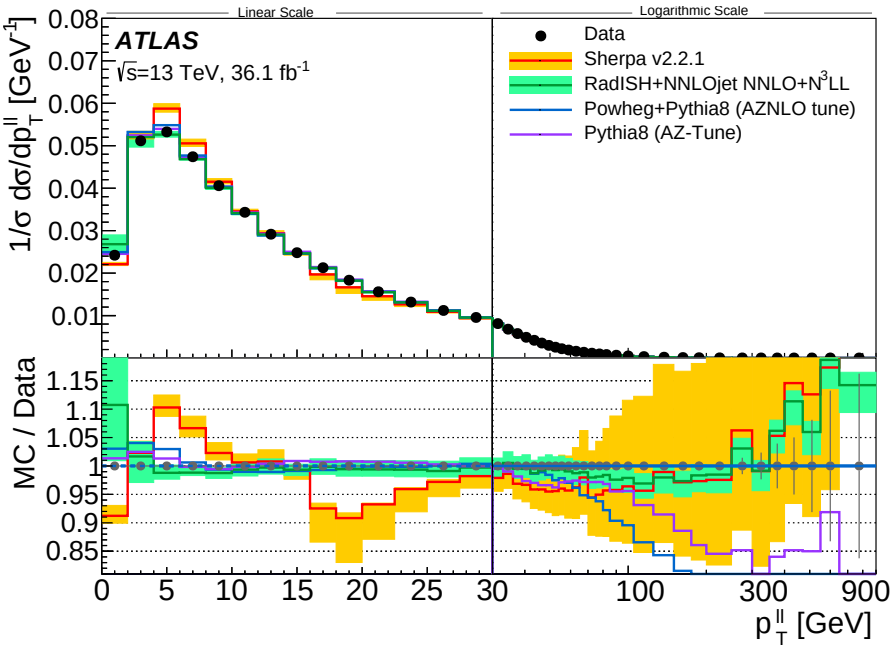


Figure 8.15: Comparison of the $p_{\mathrm{T}}^{\ell\ell}$ distributions between different MC generators and data from the ATLAS measurement [128].

In the ATLAS m_{W} measurement [122], the parton shower generator is tuned to match the $p_{\mathrm{T}}^{\mathrm{Z}}$ distribution measured in data, and the tuning is used to predict the

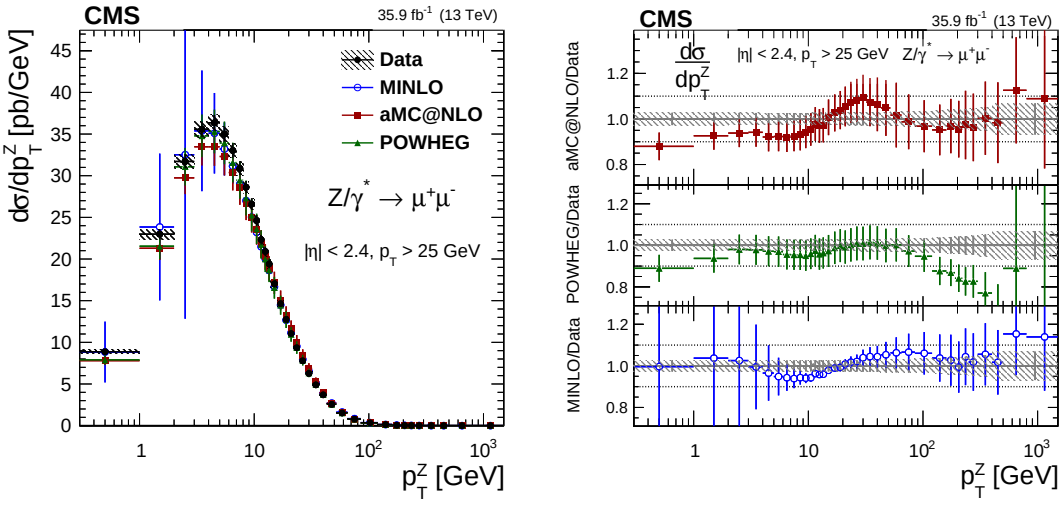


Figure 8.16: Comparison of the $p_{\mathrm{T}}^{\ell\ell}$ distributions between different MC generators and data from the CMS measurement [129].

$p_{\mathrm{T}}^{\mathrm{W}}$ distributions. Uncertainties on $p_{\mathrm{T}}^{\mathrm{W}}$ are estimated from the uncertainties in the tuning parameters. Three additional contributions to the uncertainty are coming from: the charm quark mass, the factorization scale μ_{F}, and the parton shower PDF uncertainties, each of which affects the $p_{\mathrm{T}}^{\mathrm{W}}$ and $p_{\mathrm{T}}^{\mathrm{Z}}$ distributions. Figure 8.17 shows the uncertainty on the W and Z normalized p_{T} distribution ratios relative to the nominal Pythia 8 prediction, where σ_{W} and σ_{Z} are the W and Z production cross sections.

Within the CMS Collaboration, the proper way to estimate the $p_{\mathrm{T}}^{\mathrm{W}}$ uncertainties for the m_{W} measurement is still under discussion. Here a toy model is taken, with the variation scales approximately the same as in the ATLAS analysis. While the absolute values from the toy model will not reflect the final m_{W} uncertainty from the $p_{\mathrm{T}}^{\mathrm{W}}$ modeling, it is possible to estimate the robustness of the m_{T} reconstruction against the $p_{\mathrm{T}}^{\mathrm{W}}$ modeling when different recoil estimators are employed. This method allows us to have a first comparison of the performances of these different estimators.

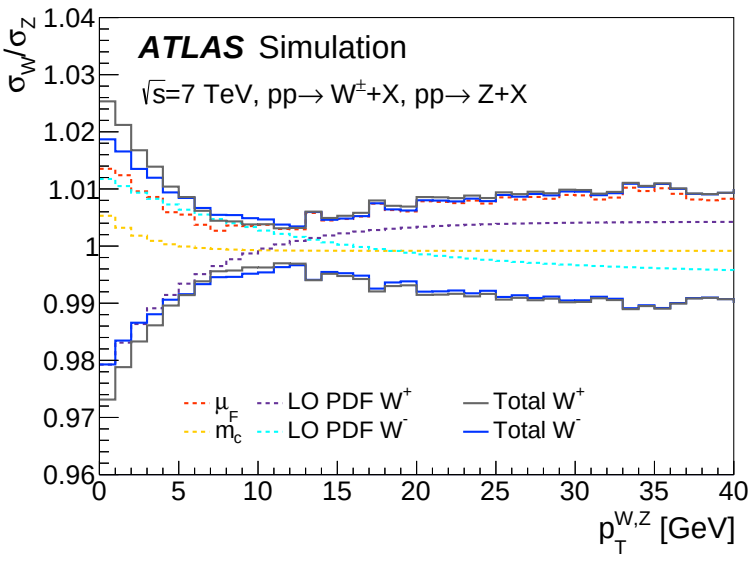


Figure 8.17: Uncertainty on the W and Z normalized transverse-momentum distribution ratios relative to the nominal Pythia 8 prediction, from the ATLAS m_{W} measurement [122].

The $p_{\mathrm{T}}^{\mathrm{W}}$ variations in Figure 8.17 from ATLAS are fitted with a third-order polynomial function. Figure 8.18 shows the fitted function, which fully covers the uncertainty band in the ATLAS measurement in Figure 8.17: the variation is about 2.5% at very low $p_{\mathrm{T}}^{\mathrm{W}}$, and about 1% at high $p_{\mathrm{T}}^{\mathrm{W}}$. For $p_{\mathrm{T}}^{\mathrm{W}} > 35\,\mathrm{GeV}$, the variation is set to a constant. This function is used to reweight the $p_{\mathrm{T}}^{\mathrm{W}}$ distributions in the MC templates.

In Figure 8.19 the default m_{T} distribution (black) is compared to the distributions obtained by shifting m_{W} by $\pm 10\,\mathrm{MeV}$ (blue and cyan), and by reweighting the $p_{\mathrm{T}}^{\mathrm{W}}$ with the above fitted function (red). Looking at the ratio distributions at the bottom of each plot, the distribution using the generator-level $p_{\mathrm{T}}^{\mathrm{W}}$ has the largest variations when shifting the m_{W} by $\pm 10\,\mathrm{MeV}$, indicating its strongest discriminating power on m_{W} and smallest statistical uncertainties. While for PF $p_{\mathrm{T}}^{\mathrm{miss}}$, the m_{T} variations from m_{W} shift is the smallest, indicating the largest statistical uncertainty

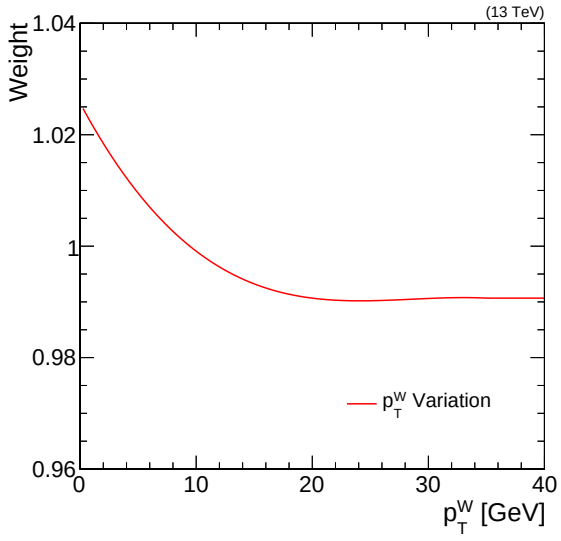
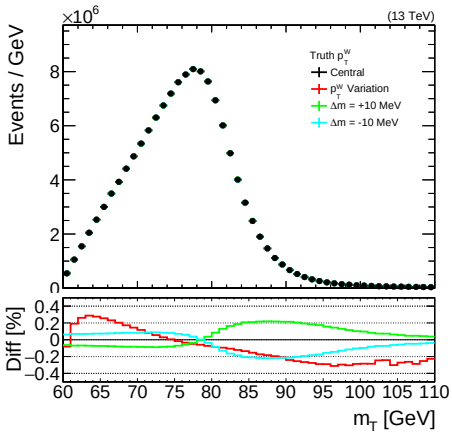


Figure 8.18: Weights for reweighting the $p_{\mathrm{T}}^{\mathrm{W}}$ spectrum.

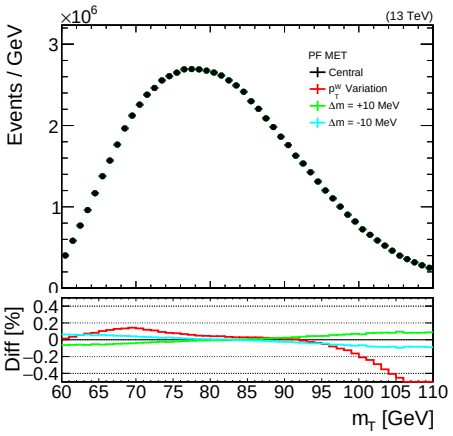
on m_{W} fit due to its poor resolution.

Comparing the variations from $p_{\mathrm{T}}^{\mathrm{W}}$ reweighting to the variations by shifting m_{W} by $\pm 10\,\mathrm{MeV}$ in Figure 8.19, the m_{T} variations reconstructed using generator-level $p_{\mathrm{T}}^{\mathrm{W}}$ seems to be the smallest, which leads to the smallest dependence on $p_{\mathrm{T}}^{\mathrm{W}}$ spectrum; the distribution which uses $u_T = 0$ seems to have the largest variations with $p_{\mathrm{T}}^{\mathrm{W}}$ reweighting, which results in the largest systematic uncertainties from $p_{\mathrm{T}}^{\mathrm{W}}$ spectrum, as expected in the discussions before.

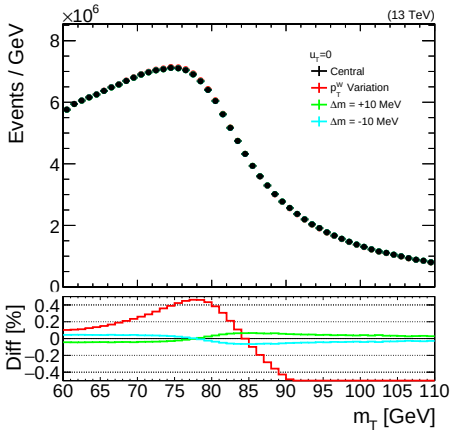
It can also be found from Figure 8.19 that the variations from $p_{\mathrm{T}}^{\mathrm{W}}$ reweighting are larger at the two ends of the m_{T} distribution than the central regions. This causes our studies to be very sensitive to the m_{T} fit range. As stated before, we take the 90% quantile as the upper bound of the m_{T} fit, to make them more comparable.



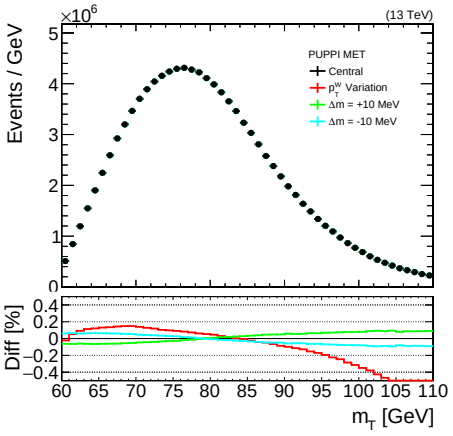
(a) Generator-level $p_{\mathrm{T}}^{\mathrm{W}}$



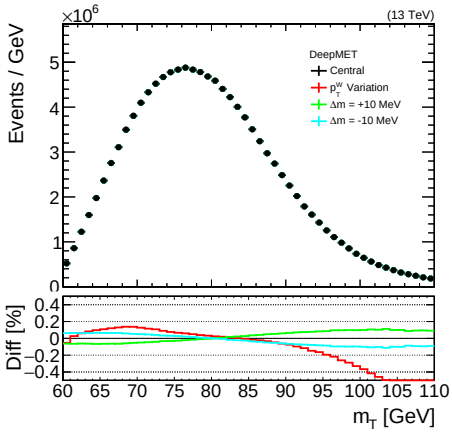
(b) PF $p_{\mathrm{T}}^{\mathrm{miss}}$



(c) $u_T = 0$



(d) PUPPI $p_{\mathrm{T}}^{\mathrm{miss}}$



(e) DeepMET

Figure 8.19: Distributions of m_{T} (black) and the variations by shifting m_{W} by 10 MeV (green and cyan), and by reweighting the $p_{\mathrm{T}}^{\mathrm{W}}$ distribution (red). The five plots from left to right, from top to bottom are the m_{T} reconstructed with the generator-level $p_{\mathrm{T}}^{\mathrm{W}}$, PF $p_{\mathrm{T}}^{\mathrm{miss}}$, $u_T = 0$, PUPPI $p_{\mathrm{T}}^{\mathrm{miss}}$ and DeepMET. The ratio plot shows the relative difference in percent between the distribution after the variation and the original (central) distribution.

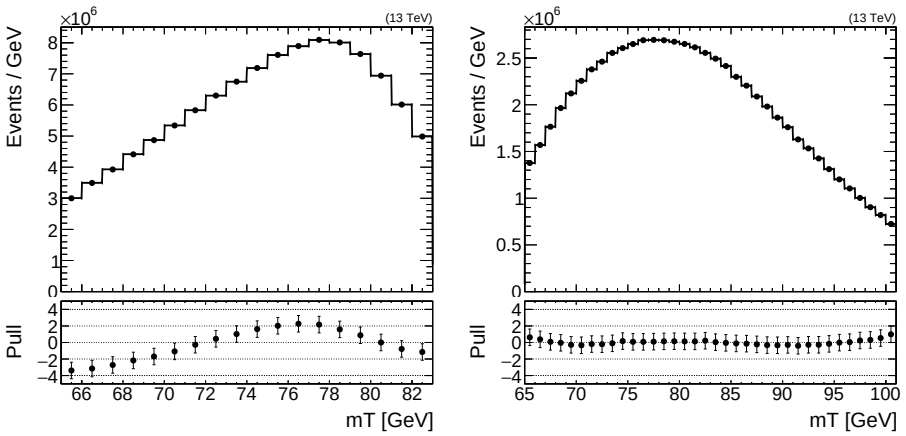
8.5.4 Uncertainties on W Mass from W pT Spectrum

With the W mass fitting procedure, event selections and m_{T} fit range, and $p_{\mathrm{T}}^{\mathrm{W}}$ variations applied as discussed above, the uncertainties on m_{W} is evaluated with different recoil estimators. The m_{T} fit is a binned maximum likelihood fit in the corresponding fit range, between the Asimov pseudo-data ($m_{\mathrm{W}} = 80419\,\mathrm{MeV}$) and the 21 morphed templates (m_{W} from 80369 MeV to 80469 MeV; one template every 5 MeV).

Without the $p_{\mathrm{T}}^{\mathrm{W}}$ variations, the fit returns 80419 MeV. After reweighting the $p_{\mathrm{T}}^{\mathrm{W}}$ distribution, the 21 templates are rebuilt from the reweighted MC, and the fit to the same pseudo-data is repeated to extract the new m_{W} values. The difference between the new m_{W} and the previous m_{W} (80419 MeV) will be taken as the estimation of the systematic uncertainties on m_{W} from the $p_{\mathrm{T}}^{\mathrm{W}}$ variations.

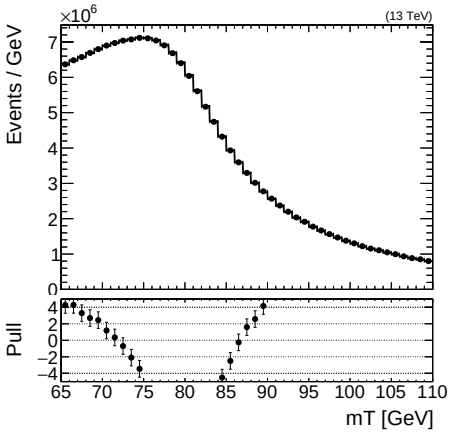
The postfit distributions are shown in Figure 8.20. The five plots are the m_{T} distributions reconstructed from the generator-level $p_{\mathrm{T}}^{\mathrm{W}}$, PF $p_{\mathrm{T}}^{\mathrm{miss}}$, $u_{\mathrm{T}} = 0$ (i.e., $m_{\mathrm{T}} = 2p_{\mathrm{T}}^{\ell}$), PUPPI $p_{\mathrm{T}}^{\mathrm{miss}}$ and DeepMET. The fit ranges are taken from Table 8.6, as discussed before. The pulls in the bottom panels are calculated as the difference between the pseudo-data and the template, divided by the statistical uncertainty of the pseudo-data. Most of the pulls after the fit are between -2 and 2, showing good agreement between the pseudo-data and the template, except for the m_{T} distribution using $u_{\mathrm{T}} = 0$, where the fitter is unable to find a good agreement within the 50 MeV fit range.

Table 8.7 shows the m_{W} statistical uncertainties and systematic uncertainties

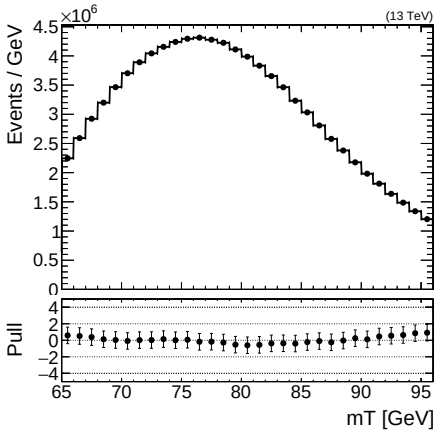


(a) Generator-level $p_{\mathrm{T}}^{\mathrm{W}}$

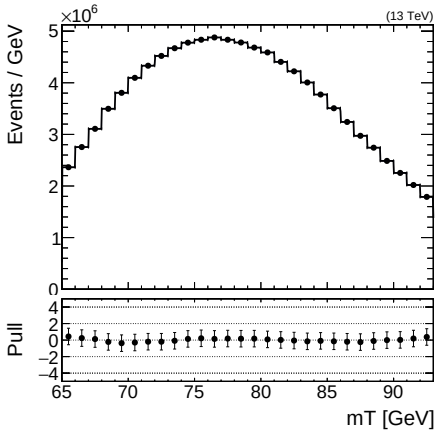
(b) PF $p_{\mathrm{T}}^{\mathrm{miss}}$



(c) $u_{\mathrm{T}} = 0$



(d) PUPPI $p_{\mathrm{T}}^{\mathrm{miss}}$



(e) DeepMET

Figure 8.20: The m_{T} postfit results between pseudo-data and morphed templates, where the morphed templates are from MC with $p_{\mathrm{T}}^{\mathrm{W}}$ reweighting applied. Pseudo-data has been scaled up to $35.9\,\mathrm{fb}^{-1}$ for the evaluation of statistical uncertainties. The five plots from left to right, from top to bottom are the m_{T} reconstructed with the generator-level $p_{\mathrm{T}}^{\mathrm{W}}$, PF $p_{\mathrm{T}}^{\mathrm{miss}}$, $u_T = 0$, PUPPI $p_{\mathrm{T}}^{\mathrm{miss}}$ and DeepMET. The plotting ranges are the same as the fitting ranges listed in Table 8.6.

from $p_{\mathrm{T}}^{\mathrm{W}}$ variations with different recoil estimators. The recoil using generator-level $p_{\mathrm{T}}^{\mathrm{W}}$ has the smallest statistical and systematic uncertainties. Using only lepton p_{T} ($u_{\mathrm{T}} = 0$) leads to small statistical uncertainties because of the excellent resolution of the muon momentum measurement, but large systematic uncertainties from the $p_{\mathrm{T}}^{\mathrm{W}}$ variations. Comparing all the reconstructed recoil estimators, DeepMET has the smallest uncertainties. The results are also shown in Figure 8.21.

Table 8.7: m_{W} statistical uncertainties and systematic uncertainties from $p_{\mathrm{T}}^{\mathrm{W}}$ variations for the m_{T} fit, from different recoil estimators. All units are in MeV.

Recoil	stat.	syst. from $p_{\mathrm{T}}^{\mathrm{W}}$ variations	total
Generator-level $p_{\mathrm{T}}^{\mathrm{W}}$	1.3	9.1	9.2
PF $p_{\mathrm{T}}^{\mathrm{miss}}$	3.9	20.3	20.7
PUPPI $p_{\mathrm{T}}^{\mathrm{miss}}$	2.6	21.4	21.6
$u_T = 0$	1.8	>50	>50
DeepMET	2.5	16.1	16.3

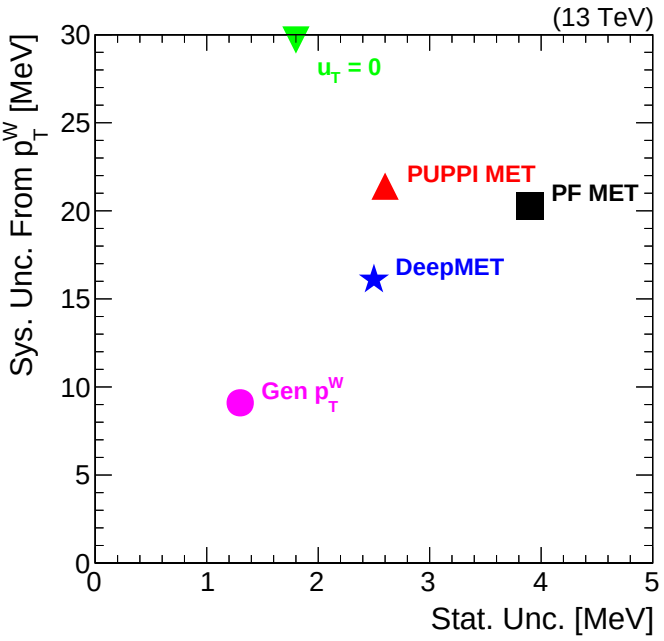


Figure 8.21: m_{W} statistical uncertainties and systematic uncertainties from $p_{\mathrm{T}}^{\mathrm{W}}$ variations for the fit of m_{T} reconstructed from different recoil estimators.

As emphasized before, the setup here to evaluate the uncertainties is rather simplified and the absolute values of the estimated uncertainties have a large dependence on the $p_{\mathrm{T}}^{\mathrm{W}}$ variations and the chosen fit range. To evaluate the final m_{W} uncertainties the methods will need further discussions and optimizations. The results from this simplified approach are however still promising: DeepMET reduces the systematic uncertainty from the $p_{\mathrm{T}}^{\mathrm{W}}$ variations by about 20% compared with the second-best estimator.

Chapter 9: Summary and Outlook

This dissertation presents the new Deep-Neural-Network–based $p_{\mathrm{T}}^{\mathrm{miss}}$ estimator, DeepMET, and its application to the measurement of the hadronic recoil against the W boson in pp collisions. DeepMET utilizes low-level reconstructed particle-flow candidates as inputs and assigns a weight w_i and two biases $b_{i,x}, b_{i,y}$ to each candidate. The estimated $p_{\mathrm{T}}^{\mathrm{miss}}$ is the negative of the vector sum of the weighted transverse momenta of all candidates plus their bias contributions. With only approximately 5000 trainable parameters, DeepMET manages to achieve 10-20% better resolutions, more resilience to pileup, and a similar unitary response, compared with the current CMS PF and PUPPI $p_{\mathrm{T}}^{\mathrm{miss}}$ estimators. Good agreement is found between the CMS 13 TeV data taken during 2016 and simulated events after implementing corrections derived from $\mathrm{Z} \to \ell\ell$ events. DeepMET demonstrates the potential to improve the precision of SM measurements and achieve a higher sensitivity in BSM searches.

The mass of the W boson is one of the most fundamental parameters in the SM. A precise measurement of the W mass provides one of the most stringent tests of the internal consistency of the SM, and reduces the parameter space of many theories beyond the SM. This dissertation studies the impact of the recoil measurement on the W mass measurement, and shows that DeepMET is capable of reducing the

uncertainties on the W recoil and W mass by about 20% relative to other $p_{\mathrm{T}}^{\mathrm{miss}}$ estimators.

DeepMET represents the first attempt of using DNN in $p_{\mathrm{T}}^{\mathrm{miss}}$ estimations. Future developments include the deployment of similar small-size DNN algorithms in FPGAs [130] in the Level-1 trigger, or in the GPUs [131] present in the high-level trigger. It is also interesting to explore more complicated and advanced ML algorithms. For instance, graph neutral networks [132] have already shown great potential in studies such as event reconstruction [133], pileup rejection [134], and jet-flavor tagging [135, 136]. The exploration and deployment of modern machine learning techniques, especially deep-learning techniques, will continue to be one of the fastest developing fields in particle physics.

Bibliography

[1] J. Thomson, "Cathode rays", *Phil. Mag. Ser. 5* **44** (1897) 293–316, doi:10.1080/14786449708621070.

[2] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett. B* **716** (2012) 1–29, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.

[3] CMS Collaboration, "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC", *Phys. Lett. B* **716** (2012) 30–61, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.

[4] "CMS Wiki page of the timescales in particle physics", 2020. Accessed: 2020-10-02. https://wiki.physik.uzh.ch/cms/latex:tikz:timescales.

[5] "Timeline of particle discoveries", 2020. Accessed: 2020-10-02. https://en.wikipedia.org/w/index.php?title=Timeline_of_particle_discoveries&oldid=977484045.

[6] CMS Collaboration, "Identification of heavy, energetic, hadronically decaying particles using machine-learning techniques", *JINST* **15** (2020), no. 06, P06005, doi:10.1088/1748-0221/15/06/P06005, arXiv:2004.08262.

[7] CMS Collaboration, "Performance of the DeepJet b tagging algorithm using 41.9/fb of data from proton-proton collisions at 13TeV with Phase 1 CMS detector", Technical Report CMS-DP-2018-058, 2018.

[8] CMS Collaboration, "A deep neural network for simultaneous estimation of b jet energy and resolution", arXiv:1912.06046.

[9] CMS Collaboration, "Performance of the DeepTau algorithm for the discrimination of taus against jets, electron, and muons", Technical Report CMS-DP-2019-033, 2019.

[10] C. Balazs and C. Yuan, "Soft gluon effects on lepton pairs at hadron colliders", *Phys. Rev. D* **56** (1997) 5558–5583, doi:10.1103/PhysRevD.56.5558, arXiv:hep-ph/9704258.

[11] M. Baak et al., "Updated Status of the Global Electroweak Fit and Constraints on New Physics", *Eur. Phys. J. C* **72** (2012) 2003, doi:10.1140/epjc/s10052-012-2003-4, arXiv:1107.0975.

[12] J. Haller et al., "Update of the global electroweak fit and constraints on two-Higgs-doublet models", *Eur. Phys. J. C* **78** (2018), no. 8, 675, doi:10.1140/epjc/s10052-018-6131-3, arXiv:1803.01853.

[13] S. Glashow, "Partial Symmetries of Weak Interactions", *Nucl. Phys.* **22** (1961) 579–588, doi:10.1016/0029-5582(61)90469-2.

[14] S. Weinberg, "A Model of Leptons", *Phys. Rev. Lett.* **19** (1967) 1264–1266, doi:10.1103/PhysRevLett.19.1264.

[15] A. Salam, "Weak and Electromagnetic Interactions", *Conf. Proc. C* **680519** (1968) 367–377, doi:10.1142/9789812795915_0034.

[16] H. Politzer, "Reliable Perturbative Results for Strong Interactions?", *Phys. Rev. Lett.* **30** (1973) 1346–1349, doi:10.1103/PhysRevLett.30.1346.

[17] D. J. Gross and F. Wilczek, "Ultraviolet Behavior of Nonabelian Gauge Theories", *Phys. Rev. Lett.* **30** (1973) 1343–1346, doi:10.1103/PhysRevLett.30.1343.

[18] A. Purcell, "Go on a particle quest at the first CERN webfest. Le premier webfest du CERN se lance á la conquête des particules", Technical Report BUL-NA-2012-269, 2012.

[19] CMS Collaboration, "Measurement of the inclusive 3-jet production differential cross section in proton–proton collisions at 7 TeV and determination of the strong coupling constant in the TeV range", *Eur. Phys. J. C* **75** (2015), no. 5, 186, doi:10.1140/epjc/s10052-015-3376-y, arXiv:1412.1633.

[20] P. W. Higgs, "Broken symmetries, massless particles and gauge fields", *Phys. Lett.* **12** (1964) 132–133, doi:10.1016/0031-9163(64)91136-9.

[21] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons", *Phys. Rev. Lett.* **13** (1964) 508–509, doi:10.1103/PhysRevLett.13.508.

[22] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons", *Phys. Rev. Lett.* **13** (1964) 321–323, doi:10.1103/PhysRevLett.13.321.

[23] G. Guralnik, C. Hagen, and T. Kibble, "Global Conservation Laws and Massless Particles", *Phys. Rev. Lett.* **13** (1964) 585–587, doi:10.1103/PhysRevLett.13.585.

[24] P. W. Higgs, "Spontaneous Symmetry Breakdown without Massless Bosons", *Phys. Rev.* **145** (1966) 1156–1163, doi:10.1103/PhysRev.145.1156.

[25] J. Ellis, "Higgs Physics", Technical Report KCL-PH-TH-2013-49. CERN-PH-TH-2013-315, Dec, 2013. doi:10.5170/CERN-2015-004.117.

[26] J. Erler and P. Langacker, "Implications of high precision experiments and the CDF top quark candidates", *Phys. Rev. D* **52** (1995) 441–450, doi:10.1103/PhysRevD.52.441, arXiv:hep-ph/9411203.

[27] M. Awramik, M. Czakon, A. Freitas, and G. Weiglein, "Precise prediction for the W boson mass in the standard model", *Phys. Rev. D* **69** (2004) 053006, doi:10.1103/PhysRevD.69.053006, arXiv:hep-ph/0311148.

[28] L. Evans and P. Bryant, "LHC Machine", *JINST* **3** (2008) S08001, doi:10.1088/1748-0221/3/08/S08001.

[29] G. Bachy et al., "The LEP collider: construction, project status and outlook", *Part. Accel.* **26** (1990) 19–32.

[30] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", *JINST* **3** (2008) S08003, doi:10.1088/1748-0221/3/08/S08003.

[31] CMS Collaboration, "The CMS Experiment at the CERN LHC", *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.

[32] ALICE Collaboration, "The ALICE experiment at the CERN LHC", *JINST* **3** (2008) S08002, doi:10.1088/1748-0221/3/08/S08002.

[33] LHCb Collaboration, "The LHCb Detector at the LHC", *JINST* **3** (2008) S08005, doi:10.1088/1748-0221/3/08/S08005.

[34] J.-L. Caron, "The LHC injection complex.. L'ensemble d'injection du LHC.", Technical Report LHC-PHO-1993-008, May, 1993. AC Collection. Legacy of AC. Pictures from 1992 to 2002.

[35] J. Blewett, "200-GeV Intersecting Storage Accelerators", *eConf* **C710920** (1971) 501.

[36] S. Dailler, "Cross section of LHC dipole. Dipole LHC: coupe transversale.", (Apr, 1999). AC Collection. Legacy of AC. Pictures from 1992 to 2002.

[37] O. S. Bruning et al., "LHC Design Report Vol.1: The LHC Main Ring", doi:10.5170/CERN-2004-003-V-1.

[38] J. Wenninger, "Operation and Configuration of the LHC in Run 2", Technical Report CERN-ACC-NOTE-2019-0007, Mar, 2019.

[39] "CMS Luminosity Public Results". https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults. Accessed: 2020-09-15.

[40] CMS Collaboration, "Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV", *JHEP* **07** (2018) 161, doi:10.1007/JHEP07(2018)161, arXiv:1802.02613.

[41] ATLAS Collaboration, "Measurement of the Inelastic Proton-Proton Cross Section at $\sqrt{s} = 13$ TeV with the ATLAS Detector at the LHC", *Phys. Rev. Lett.* **117** (2016), no. 18, 182002, doi:10.1103/PhysRevLett.117.182002, arXiv:1606.02625.

[42] CMS Collaboration, "Cutaway diagrams of CMS detector", Technical Report CMS-OUTREACH-2019-001, May, 2019.

[43] CMS Collaboration, "Precise Mapping of the Magnetic Field in the CMS Barrel Yoke using Cosmic Rays", *JINST* **5** (2010) T03021, doi:10.1088/1748-0221/5/03/T03021, arXiv:0910.5530.

[44] "CMS Detector Layout". http://www.hephy.at/user/friedl/diss/html/node8.html. Accessed: 2020-09-21.

[45] "CMS Wiki page of the coordinate system". https://wiki.physik.uzh.ch/cms/latex:example_spherical_coordinates. Accessed: 2020-09-17.

[46] CMS Collaboration, "The CMS tracker system project: Technical Design Report", Technical Report CERN-LHCC-98-006, CMS-TDR-5, 1997.

[47] CMS Collaboration, "The CMS tracker: addendum to the Technical Design Report", Technical Report CERN-LHCC-2000-016, CMS-TDR-5-add-1, 2000.

[48] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *JINST* **9** (2014), no. 10, P10009, doi:10.1088/1748-0221/9/10/P10009, arXiv:1405.6569.

[49] CMS Collaboration, "Phase 1 upgrade of the CMS pixel detector", *JINST* **12** (2017), no. 02, C02033, doi:10.1088/1748-0221/12/02/C02033.

[50] CMS Collaboration, "Status of the CMS Phase I Pixel Detector Upgrade", *Nucl. Instrum. Meth. A* **831** (2016) 71–75, doi:10.1016/j.nima.2016.03.028, arXiv:1511.06084.

[51] CMS Collaboration, "The CMS electromagnetic calorimeter project: Technical Design Report", Technical Report CERN-LHCC-97-033, CMS-TDR-4, 1997.

[52] CMS Collaboration, "Energy Calibration and Resolution of the CMS Electromagnetic Calorimeter in pp Collisions at $\sqrt{s} = 7\,\text{TeV}$", *JINST* **8** (2013) 9009, doi:10.1088/1748-0221/8/09/P09009, arXiv:1306.2016.

[53] P. Adzic et al., "Energy resolution of the barrel of the CMS electromagnetic calorimeter", *JINST* **2** (2007) P04004, doi:10.1088/1748-0221/2/04/P04004.

[54] CMS Collaboration, "CMS: The hadron calorimeter technical design report", Technical Report CERN-LHCC-97-031, CMS-TDR-4, 1997.

[55] CMS Collaboration, "Calibration of the CMS hadron calorimeters using proton-proton collision data at $\sqrt{s} = 13\,\text{TeV}$", *JINST* **15** (2020), no. 05, P05002, doi:10.1088/1748-0221/15/05/P05002, arXiv:1910.00079.

[56] CMS HCAL Collaboration, "Design, performance, and calibration of CMS hadron-barrel calorimeter wedges", *Eur. Phys. J. C* **55** (2008), no. 1, 159–171, doi:10.1140/epjc/s10052-008-0573-y.

[57] CMS Collaboration, "CMS Technical Design Report for the Phase 1 Upgrade of the Hadron Calorimeter", Technical Report CERN-LHCC-2012-015, CMS-TDR-010, 2012. doi:10.2172/1151651.

[58] CMS Collaboration, "The CMS muon project: Technical Design Report", Technical Report CERN-LHCC-97-032, CMS-TDR-3, 1997.

[59] CMS Collaboration, "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV", *JINST* **13** (2018), no. 06, P06015, doi:10.1088/1748-0221/13/06/P06015, arXiv:1804.04528.

[60] CMS Collaboration, "The CMS trigger system", *JINST* **12** (2017), no. 01, P01020, doi:10.1088/1748-0221/12/01/P01020, arXiv:1609.02366.

[61] CMS Collaboration, "CMS. The TriDAS project. Technical design report, vol. 1: The trigger systems", Technical Report CERN-LHCC-2000-038, CMS-TDR-6-1, 2000.

[62] CMS Collaboration, "CMS: The TriDAS project. Technical design report, Vol. 2: Data acquisition and high-level trigger", Technical Report CERN-LHCC-2002-026, CMS-TDR-6, 2002.

[63] CMS Collaboration, "Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13$ TeV", (2020). arXiv:2006.10165. Submitted to *JINST*.

[64] CMS Collaboration Collaboration, "Data Parking and Data Scouting at the CMS Experiment", Technical Report CMS-DP-2012-022, 2012.

[65] CMS Collaboration, "Interactive Slice of the CMS detector", Technical Report CMS-OUTREACH-2016-027, Aug, 2016.

[66] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector", *JINST* **12** (2017), no. 10, P10003, doi:10.1088/1748-0221/12/10/P10003, arXiv:1706.04965.

[67] D. Bertolini, P. Harris, M. Low, and N. Tran, "Pileup Per Particle Identification", *JHEP* **10** (2014) 059, doi:10.1007/JHEP10(2014)059, arXiv:1407.6013.

[68] R. Frühwirth, "Application of Kalman filtering to track and vertex fitting", *Nucl. Instrum. Meth. A* **262** (1987) 444–450, doi:10.1016/0168-9002(87)90887-4.

[69] K. Rose, "Deterministic annealing for clustering, compression, classification, regression, and related optimization problems", *IEEE Proc.* **86** (1998), no. 11, 2210–2239, doi:10.1109/5.726788.

[70] R. Frühwirth, W. Waltenberger, and P. Vanlaer, "Adaptive vertex fitting", *J. Phys. G* **34** (2007) N343, doi:10.1088/0954-3899/34/12/N01.

[71] CMS Collaboration, "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV", *JINST* **13** (2018), no. 06, P06015, doi:10.1088/1748-0221/13/06/P06015, arXiv:1804.04528.

[72] CMS Collaboration, "Muon Reconstruction and Identification Performance with Run-2 data", Technical Report CMS-DP-2020-040, 2020.

[73] CMS Collaboration, "Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV", *JINST* **10** (2015), no. 06, P06005, doi:10.1088/1748-0221/10/06/P06005, arXiv:1502.02701.

[74] CMS Collaboration, "Performance of Photon Reconstruction and Identification with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV", *JINST* **10** (2015), no. 08, P08010, doi:10.1088/1748-0221/10/08/P08010, arXiv:1502.02702.

[75] CMS Collaboration, "Jet algorithms performance in 13 TeV data", Technical Report CMS-PAS-JME-16-003, CERN, Geneva, 2017.

[76] CMS Collaboration, "Pileup mitigation at CMS in 13 TeV data", (2020). arXiv:2003.00503. Submitted to *JINST*.

[77] G. P. Salam, "Towards Jetography", *Eur. Phys. J. C* **67** (2010) 637–686, doi:10.1140/epjc/s10052-010-1314-6, arXiv:0906.1833.

[78] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm", *JHEP* **04** (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.

[79] S. D. Ellis and D. E. Soper, "Successive combination jet algorithm for hadron collisions", *Phys. Rev. D* **48** (1993) 3160–3166, doi:10.1103/PhysRevD.48.3160, arXiv:hep-ph/9305266.

[80] Y. L. Dokshitzer, G. Leder, S. Moretti, and B. Webber, "Better jet clustering algorithms", *JHEP* **08** (1997) 001, doi:10.1088/1126-6708/1997/08/001, arXiv:hep-ph/9707323.

[81] M. Wobisch and T. Wengler, "Hadronization corrections to jet cross-sections in deep inelastic scattering", in *Workshop on Monte Carlo Generators for HERA Physics (Plenary Starting Meeting)*, pp. 270–279. 4, 1998. arXiv:hep-ph/9907280.

[82] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet User Manual", *Eur. Phys. J. C* **72** (2012) 1896, doi:10.1140/epjc/s10052-012-1896-2, arXiv:1111.6097.

[83] CMS Collaboration, "Jet energy scale and resolution performance with 13 TeV data collected by CMS in 2016-2018", Technical Report CMS-DP-2020-019, 2020.

[84] CMS Collaboration, "Missing transverse energy performance of the CMS detector", *JINST* **6** (2011) P09001, doi:10.1088/1748-0221/6/09/P09001, arXiv:1106.5048.

[85] CMS Collaboration, "Performance of the CMS missing transverse momentum reconstruction in pp data at $\sqrt{s} = 8$ TeV", *JINST* **10** (2015), no. 02, P02006, doi:10.1088/1748-0221/10/02/P02006, arXiv:1411.0511.

[86] CMS Collaboration, "Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector", *JINST* **14** (2019), no. 07, P07004, doi:10.1088/1748-0221/14/07/P07004, arXiv:1903.06078.

[87] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning", *Nature* **521** (2015), no. 7553, 436–444, doi:10.1038/nature14539.

[88] I. Goodfellow, Y. Bengio, and A. Courville, "Deep Learning". MIT Press, 2016. http://www.deeplearningbook.org.

[89] A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola, "Dive into Deep Learning". 2019. http://www.d2l.ai.

[90] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines", in *ICML*, pp. 807–814. 2010.

[91] "Universal approximation theorem", 2020. Accessed: 2020-11-08. https://en.wikipedia.org/w/index.php?title=Universal_approximation_theorem&oldid=984220892.

[92] L. N. Smith, "A disciplined approach to neural network hyper-parameters: Part 1 – learning rate, batch size, momentum, and weight decay", arXiv:1803.09820.

[93] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift", arXiv:1502.03167.

[94] "Embedding layer in Keras API", 2020. Accessed: 2020-11-08. https://keras.io/api/layers/core_layers/embedding/.

[95] J. Alwall et al., "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", *JHEP* **07** (2014) 079, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.

[96] P. Nason, "A New method for combining NLO QCD with shower Monte Carlo algorithms", *JHEP* **11** (2004) 040, doi:10.1088/1126-6708/2004/11/040, arXiv:hep-ph/0409146.

[97] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method", *JHEP* **11** (2007) 070, doi:10.1088/1126-6708/2007/11/070, arXiv:0709.2092.

[98] S. Alioli, P. Nason, C. Oleari, and E. Re, "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", *JHEP* **06** (2010) 043, doi:10.1007/JHEP06(2010)043, arXiv:1002.2581.

[99] T. Sjstrand et al., "An introduction to PYTHIA 8.2", *Comput. Phys. Commun.* **191** (2015) 159–177, doi:10.1016/j.cpc.2015.01.024, arXiv:1410.3012.

[100] CMS Collaboration, "Event generator tunes obtained from underlying event and multiparton scattering measurements", *Eur. Phys. J. C* **76** (2016), no. 3, 155, doi:10.1140/epjc/s10052-016-3988-x, arXiv:1512.00815.

[101] M. L. Mangano, M. Moretti, F. Piccinini, and M. Treccani, "Matching matrix elements and shower evolution for top-quark production in hadronic collisions", *JHEP* **01** (2007) 013, doi:10.1088/1126-6708/2007/01/013, arXiv:hep-ph/0611129.

[102] R. Frederix and S. Frixione, "Merging meets matching in MC@NLO", *JHEP* **12** (2012) 061, doi:10.1007/JHEP12(2012)061, arXiv:1209.6215.

[103] NNPDF Collaboration, "Parton distributions for the LHC Run II", *JHEP* **04** (2015) 040, doi:10.1007/JHEP04(2015)040, arXiv:1410.8849.

[104] Geant4 Collaboration, "Geant4 toolkit for simulation of HEP experiments", *Nucl. Instrum. Meth. A* **502** (2003) 666–668, doi:10.1016/S0168-9002(03)00538-2.

[105] T. P. Minka, "Automatic choice of dimensionality for PCA", in *Advances in neural information processing systems*, pp. 598–604. 2001.

[106] F. Pedregosa et al., "Scikit-learn: Machine learning in Python", *Journal of Machine Learning Research* **12** (2011) 2825–2830.

[107] F. Chollet et al., "Keras". https://keras.io, 2015.

[108] M. Abadi et al., "TensorFlow: Large-scale machine learning on heterogeneous systems", 2015. Software available from tensorflow.org. https://www.tensorflow.org/.

[109] K. S. Cranmer, "Kernel estimation in high-energy physics", *Comput. Phys. Commun.* **136** (2001) 198–207, doi:10.1016/S0010-4655(00)00243-5, arXiv:hep-ex/0011057.

[110] UA1 Collaboration, "Intermediate Vector Boson Properties at the CERN Super Proton Synchrotron Collider", *Europhys. Lett.* **1** (1986) 327–345, doi:10.1209/0295-5075/1/7/002.

[111] UA2 Collaboration, "An Improved determination of the ratio of W and Z masses at the CERN $\bar{p}p$ collider", *Phys. Lett. B* **276** (1992) 354–364, doi:10.1016/0370-2693(92)90332-X.

[112] CDF Collaboration, "Measurement of the W boson mass with the Collider Detector at Fermilab", *Phys. Rev. D* **64** (2001) 052001, doi:10.1103/PhysRevD.64.052001, arXiv:hep-ex/0007044.

[113] D0 Collaboration, "Improved W Boson Mass Measurement with the D0 Detector", *Phys. Rev. D* **66** (2002) 012001, doi:10.1103/PhysRevD.66.012001, arXiv:hep-ex/0204014.

[114] CDF, D0 Collaboration, "Combination of CDF and D0 Results on W Boson Mass and Width", *Phys. Rev. D* **70** (2004) 092008, doi:10.1103/PhysRevD.70.092008, arXiv:hep-ex/0311039.

[115] CDF Collaboration, "Precise measurement of the W-boson mass with the CDF II detector", *Phys. Rev. Lett.* **108** (2012) 151803, doi:10.1103/PhysRevLett.108.151803, arXiv:1203.0275.

[116] D0 Collaboration, "Measurement of the W Boson Mass with the D0 Detector", *Phys. Rev. Lett.* **108** (2012) 151804, doi:10.1103/PhysRevLett.108.151804, arXiv:1203.0293.

[117] CDF, D0 Collaboration, "Combination of CDF and D0 W-Boson Mass Measurements", *Phys. Rev. D* **88** (2013), no. 5, 052018, doi:10.1103/PhysRevD.88.052018, arXiv:1307.7627.

[118] ALEPH Collaboration, "Measurement of the W boson mass and width in e^+e^- collisions at LEP", *Eur. Phys. J. C* **47** (2006) 309–335, doi:10.1140/epjc/s2006-02576-8, arXiv:hep-ex/0605011.

[119] DELPHI Collaboration, "Measurement of the Mass and Width of the W Boson in e^+e^- Collisions at $\sqrt{s} = 161$-209GeV", *Eur. Phys. J. C* **55** (2008) 1–38, doi:10.1140/epjc/s10052-008-0585-7, arXiv:0803.2534.

[120] L3 Collaboration, "Measurement of the mass and the width of the W boson at LEP", *Eur. Phys. J. C* **45** (2006) 569–587, doi:10.1140/epjc/s2005-02459-6, arXiv:hep-ex/0511049.

[121] OPAL Collaboration, "Measurement of the mass and width of the W boson", *Eur. Phys. J. C* **45** (2006) 307–335, doi:10.1140/epjc/s2005-02440-5, arXiv:hep-ex/0508060.

[122] ATLAS Collaboration, "Measurement of the W-boson mass in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector", *Eur. Phys. J. C* **78** (2018), no. 2, 110, doi:10.1140/epjc/s10052-017-5475-4, arXiv:1701.07240. [Erratum: Eur.Phys.J.C 78, 898 (2018)].

[123] J. Smith, W. van Neerven, and J. Vermaseren, "The Transverse Mass and Width of the W Boson", *Phys. Rev. Lett.* **50** (1983) 1738, doi:10.1103/PhysRevLett.50.1738.

[124] V. Buge et al., "Prospects for the precision measurement of the W mass with the CMS detector at the LHC", *J. Phys. G* **34** (2007) N193–N220, doi:10.1088/0954-3899/34/5/N02.

[125] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics", *Eur. Phys. J. C* **71** (2011) 1554, doi:10.1140/epjc/s10052-011-1554-0, arXiv:1007.1727. [Erratum: Eur.Phys.J.C 73, 2501 (2013)].

[126] O. Cerri, "Hadronic recoil in the W boson production at LHC for a W mass measurement with the CMS experiment", Master's thesis, Pisa U., 2017. Presented 20 Sep 2017.

[127] N. Foppiani, "Definition and calibration of the hadronic recoil in view of the measurement of the W boson mass with the CMS experiment", Master's thesis, Pisa U., 2017. Presented 20 Sep 2017.

[128] ATLAS Collaboration, "Measurement of the transverse momentum distribution of Drell–Yan lepton pairs in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *Eur. Phys. J. C* **80** (2020), no. 7, 616, doi:10.1140/epjc/s10052-020-8001-z, arXiv:1912.02844.

[129] CMS Collaboration, "Measurements of differential Z boson production cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV", *JHEP* **12** (2019) 061, doi:10.1007/JHEP12(2019)061, arXiv:1909.04133.

[130] J. Duarte et al., "FPGA-accelerated machine learning inference as a service for particle physics computing", *Comput. Softw. Big Sci.* **3** (2019), no. 1, 13, doi:10.1007/s41781-019-0027-2, arXiv:1904.08986.

[131] J. Krupa et al., "GPU coprocessors as a service for deep learning inference in high energy physics", arXiv:2007.10359.

[132] Z. Wu et al., "A Comprehensive Survey on Graph Neural Networks", *IEEE Transactions on Neural Networks and Learning Systems* (2020) 121, doi:10.1109/tnnls.2020.2978386, arXiv:1901.00596.

[133] X. Ju et al., "Graph Neural Networks for Particle Reconstruction in High Energy Physics detectors", in *33rd Annual Conference on Neural Information Processing Systems*. 3, 2020. arXiv:2003.11603.

[134] J. Arjona Martínez et al., "Pileup mitigation at the Large Hadron Collider with graph neural networks", *Eur. Phys. J. Plus* **134** (2019), no. 7, 333, doi:10.1140/epjp/i2019-12710-3, arXiv:1810.07988.

[135] E. A. Moreno et al., "JEDI-net: a jet identification algorithm based on interaction networks", *Eur. Phys. J. C* **80** (2020), no. 1, 58, doi:10.1140/epjc/s10052-020-7608-4, arXiv:1908.05318.

[136] H. Qu and L. Gouskos, "ParticleNet: Jet Tagging via Particle Clouds", *Phys. Rev. D* **101** (2020), no. 5, 056019, doi:10.1103/PhysRevD.101.056019, arXiv:1902.08570.